Towards a Four Factor Theory of Anticipatory Learning

Mark Witkowski

Interactive and Intelligent Systems Section
Department of Electrical &Electronic Engineering
Imperial College
Exhibition Road
London SW7 2BT
United Kingdom
{m.witkowski@imperial.ac.uk}

Abstract. This paper takes an overtly anticipatory stance to the understanding of animat learning and behavior. It analyses four major animal learning theories and attempts to identify the anticipatory and predictive elements inherent to them, and to provide a new unifying approach based on the anticipatory nature of those elements based on five simple predictive "rules". These rules encapsulate all the principal properties of the four diverse theories (the four factors) and provide a simple framework for understanding how an individual animat may appear to operate according to different principles under varying circumstances. The paper then indicates how these anticipatory principles can be used to define a more detailed set of postulates for the Dynamic Expectancy Model of animat learning and behavior, and to construct its computer implementation SRS/E. Some of the issues discussed are illustrated with an example experimental procedure using SRS/E.

1 Introduction

This paper takes a particular stance on animat behavior generation and learning. At the heart of this problem is how an animat should select actions to perform, under what conditions and to what purpose. It will argue that the generators of animat behavior have a strong anticipatory or predictive quality, and that learning, and our animal models of learning, should exploit the anticipatory and predictive properties inherent in an animat's structure. The ability of entities, including living organisms and machines, to anticipate future events and be in a position to react to them in a timely manner has long been recognised as a key attribute of intelligence For instance, the discussions between Charles Babbage and Italian scientists in 1840, where the meeting concluded that "… *intelligence would be measured by the capacity for anticipation*" [15].

Recently, a growing number of researchers have identified and emphasized the importance of anticipation as the basis of models of animal learning and behavior. Butz, Sigaud and Gerard [9] categorize four main distinctions between different forms

of anticipatory learning and behavior: implicit anticipation, payoff anticipation, sensorial anticipation and state anticipation. This paper focuses on the last of these, systems that exploit the properties of anticipation between states or partial state descriptions. Stolzmann *et al.* [18] describe a classifier system model based on anticipatory principles, Tani and Nolfi [20] an Artificial Neural Network approach and Witkowski the Dynamic Expectancy Model (DEM, [25], [26], [27]), which places anticipation and prediction at the center of the learning and behavioral process. The anticipatory stance imbues an animat with several important properties. First, the ability to determine possible future situations following on from the current one, thereby anticipating those situations that might be advantageous (or harmful) in anticipation of them occurring. Second, to determine the possible outcome of actions made by the animat, leading directly to the ability to establish chains or plans of actions to satisfy some desired outcome. Third, the ability to rank the effectiveness of each action in its immediate context, independently of any particular goal or task specific reward or reinforcement (by "corroboration"). Fourth, to determine when structural learning should take place, by detecting when unpredicted events occur.

From time to time goals, activities required of the animat, will arise. By constructing a Dynamic Policy Map (DPM), a graph like structure derived from the predictions it has discovered during its lifespan and then determining an intersection of this graph with the goals and current circumstances, the Dynamic Expectancy Model may determine appropriate actions to satisfy those goals. Part of the structure of the DEM provides the animat with rules by which this discovery process proceeds. Part imbues the animat with sufficient behavior to set goals and to initiate and continue all these activities until learned behavior may take over from the innate.

Section two of this paper reviews four well-established theories of animal learning: behaviorist, classical conditioning, operant (or instrumental) conditioning and cognitive or 'expectancy' models. During the 20th century each attracted strong proponents and equally strong opponents, and each was the dominant theory for a time. Each position is supported by (typically large numbers of) detailed and fully repeatable experiments. However, none of these stances could be made to explain the full range of observable behaviors, and none was able to gain an overall dominance of the others. Yet the fact remains that each regime can be shown to be present in a single animal (though not all animal species will necessarily demonstrate every attribute). Each is made manifest in the animal according to the experimental procedures to which it is subjected. Examples will be drawn from both the animal and artificial animat research domains.

Section three analyses (selected) data from each school with the express purpose of generating a new, unifying, set of principles or "rules" of prediction and propagation, specifically related to the anticipatory properties that can be extracted from the observations leading to the four models of learning and behavior. This section also reviews a number of computer models inspired by each of these the four stances.

These rules are presented and discussed in section four. The purpose of this section is to consider the anticipatory role of prediction as a unifying factor between these approaches to learning, where previously differences may have been emphasized. This section introduces the primary contribution of this paper. In developing the unifying, anticipatory, framework, this paper does not suggest that any of these theories are in any sense incorrect, only that they each need to be viewed in the context of the whole animal, and of each other, to provide a satisfying explanation of

the role of each part. Equally, no assertion is made that, either taken individually or as a whole, these theories represent a complete description of the perceptual, behavioral or learning capabilities of any individual or type of animal.

Section five develops these arguments to show how they have influenced the development of the Dynamic Expectancy Model. Dynamic Expectancy Model animats may be seen as machines for devising hypotheses that make predictions about future events, conducting experiments to corroborate them and subsequently using the knowledge they have gained to perform useful behaviors. A critical feature is the creation and corroboration of these self-testing experiments, each derived from simple "micro-hypotheses", which are in turn created directly from observations in the environment. Each hypothesis will be viewed as describing and encapsulating a simple experiment. Each "micro-experiment" takes the form of an expectancy or prediction that is either fulfilled, so corroborating the effectiveness of the hypothesis, or is not fulfilled, weakening or denying the hypothesis. Anticipatory principles seem interesting in this context as they define a continuing process of discovery and refinement. This allows an animat to progress throughout its lifetime; incrementally developing is structures, and so match its behavior patterns to its environment.

Section six briefly describes the control architecture for SRS/E, an implementation of the Dynamic Expectancy Model. Section seven presents some illustrative results from an experimental procedure with the model. Further results using this model have been previously reported elsewhere ([25], [26], [27]).

2 Prediction and Theories of Behavior

We continue with the view that behavior generation ("action selection") is properly described by the direct or indirect interaction of sensed conditions, Sign-stimuli (S) and response, action or behavior (R) generators. This section will outline four major theoretical stances relating to animal behavior and learning, and will particularly focus on those issues relating to predictive ability, which will be considered in further detail later.

2.1 The Behaviorist Approach

It has been a long established and widely held truism that much of the behavior observed in natural animals can be described in terms of actions initiated by the current conditions in which the animal finds itself. This approach has a long tradition in the form of stimulus-response (S-R) behaviorism, and, although proposed over a century ago ([22]), continues to find proponents, for instance in the behavior based models of Maes [11], the reactive or situated models of Agre [1] and Bryson [8], and was a position vigorously upheld by Brooks [7] in his "intelligence without reason" arguments.

All argue that the majority of observed and apparently intelligent behavior may be ascribed to an innate, pre-programmed, stimulus response mechanism available to the individual. Innate intelligence is not, however, defined by degree. Complex, essentially reactive, models have been developed to comprehensively describe and (so

largely) explain the behavioral repertoire of several non-primate vertebrate species, including small mammals, birds and fish. Tyrrell [24] provides a useful summary of a variety of action selection mechanisms drawn from natural and artificial examples.

Behaviorist learning is considered to be "reinforcement", or strengthening of the activating bond between stimulus and response. That is the occurrence of a desirable event concurrently (or immediately following) an application of the S-R pair enhances the likelihood that the pairing will be invoked again over other, alternative pairings, conversely, with a reduced likelihood for undesirable events. New pairings may be established by creating an S-R link between a stimulus and a response that were active concurrently with (or immediately preceding) the desired event.

2.2 Classical Conditioning

A second, deeply influential, approach to animal learning developed during the 1920's as a result of the work of Ivan Pavlov (1849-1936), now usually referred to as classical conditioning. The procedure is well known and highly repeatable. It is neatly encapsulated by one of the earliest descriptions provided by Pavlov. Dogs naturally salivate in response to the smell or taste of meat powder. Salivation is the *unconditioned reflex* (UR), instigated by the *unconditioned stimulus* (US), the meat powder. Normally the sound of a bell does not cause the animal to salivate. If the bell is sounded almost simultaneously with the presentation of meat powder over a number of trials, it is subsequently found that the sound of the bell alone will cause salivation. The sound has become a *conditioned stimulus* (CS). The phenomena is widespread, leading Bower and Hilgard ([6], p. 58) to comment "*almost anything that moves, squirts or wiggles could be conditioned if a response from it can be reliably and repeatably evoked by a controllable unconditioned stimulus*".

The conditioned response develops with a characteristic sigmoid curve with repeated CS/US pairings. Once established the CS/UR pairing will diminish if the CS/US pairing is not regularly maintained (*extinction*). We may note that the scope of the US may be manipulated over a number of trials to either be highly differentiated to a specific signal, or conversely gradually generalized to respond to a range of similar signals (for instance, a tone of particular frequency, versus a range of frequencies about a center). *Higher-order conditioning* ([3]; [6], p. 62) allows a second neutral CS' (say, a light) to be conditioned to an existing CS (the bell), using the standard procedure. CS' then elicits the CR.

2.3 Operant Conditioning

An radically alternative view of learning was proposed by B.F. Skinner (1904-1990), that of instrumental or operant conditioning. In this model, responses are not "elicited" by impinging sensory conditions, but "emitted" by the animal in anticipation of a reward outcome. Reinforcement strengthening is therefore considered to be between response (R) and rewarding outcome (O), the R-O model, not between sensation and action, as in the S-R model.

The approach is illustrated by reference to an experimental apparatus developed by Skinner to test the paradigm, now universally referred to as the "Skinner box". In a

typical Skinner box the subject animal (typically a rat) operates a lever to obtain a reward, say a small food pellet. The subject must be prepared by the experimenter to associate operating the lever with the food reward. However, once the subject is conditioned in this manner the apparatus may be used to establish various regimes to investigate effects such as stimulus differentiation, experimental extinction, the effects of adverse stimuli ("punishment schedules") and the effects of different schedules of reinforcement (such as varying the frequency of reward). As the apparatus may be set up to automatically record the activities of the subject animal (lever pressing), long and/or complicated schedules are easily established.

Operant conditioning has found application in behavior "shaping" techniques, where an experimenter wishes to directly manipulate the overt behavioral activities of a subject, animal or human. In the simplest case the experimenter waits for the subject to emit the desired behavior, and immediately afterwards presents a reward (before a rat may be used in a Skinner box it must be prepared in this way). Importantly, it is to be noted that the R-O activity may be easily manipulated so as to occur only in the presence of a specific stimulus, which may in turn be differentiated or generalized by careful presentation or withholding of reward in the required circumstances.

This has lead to the assertion that operant conditioning is properly described by as three-part association, S-R-O. It is also interesting to note that the stimulus itself now appears to act as a conditioned reinforcer, where it had no inherent reinforcing properties before. In turn, then, a new response in the context of another stimulus (Sy) and response (Ry) may be conditioned to the existing triple (Sx-Rx-O):

Sy-Ry-Sx-Rx-O

Chains of considerable length and complexity have been generated in this way, and have been used, for instance, in the film industry to prepare performing animals. It is, of course, a given that the rewarding outcome is itself a sensory event with direct (innate) association with some condition the subject wants (or in the case of aversive condition, does not want). If the subject animal is not, for instance, hungry when offered food, the connection will not be manifest, and might not be formed. It is also the case that an apparently non-reinforcing sensory condition can attain reinforcing properties if presented in conjunction with an innately reinforcing (positive or negative) one, the *secondary* or *derived reinforcement* effect ([6], p. 184). Derived reinforcers will also condition responses unrelated to the original one.

2.4 The "Cognitive" Model:

In the final model to be considered, derived from Tolman's [23] notion of a *Sign-Gestalt Expectancy*, that is a three part "basic cognitive unit" of the form S1-R-S2, in which the occurrence of the stimulus S1 in conjunction with the activity R, leads to the expectation or prediction of the outcome S2 (which may or may not be "rewarding"). This is largely equivalent to Catania's [10] description of the fully discriminated operant connection as a *three-part contingency* of "stimulus – response – consequence", but with the essential difference that it is the identity of the outcome that is to be recorded, rather than just a measure of the desirability or quality of the connection as assumed in purely behaviorist S-R or operant conditioning approaches.

Tolman's means-ends approach inspired, and remains one of the central techniques of, Artificial Intelligence problem solving and planning techniques.

3 Interpreting Behavior as Prediction

It is clear that the standard S-R formulation makes no explicit prediction as to the outcome of performing the action part. But there is nevertheless an implicit prediction that responding in this way will place the animal in a "better" situation than the current one, and that the animal will be driven forward to a situation where further behaviors are triggered. Maes' model [11] makes this explicit. The S-R model is an effective one, and explains much about innate behavior generation. However the implicit prediction is one shared with the species as a whole (actually with the forebears of the individual).

Modern reinforcement learning techniques ([19], for a recent review) have revitalized our view of how this implicit prediction should be viewed. They provide robust and analytically tractable ways to guarantee the prioritization of multiple S-R connections to achieve optimized performance. Such policy maps, while finding many important applications, tend to be "over stable" with respect to sources of reward. In contrast, when reward states change, animals respond quickly to these changing needs.

The anticipatory attributes of the classical conditioning paradigm have long been noted, not least because it is almost impossible to establish the effect when the CS occurs after the US. Indeed for best results the CS must be presented (a short time) before the US, implying that there is a predictive effect. It remains an open question as to whether classical conditioning should be interpreted as a general predictive principle, or if it is just a highly specific phenomenon only associated with autonomic reflexes. This paper tends on the side of generality. Classical conditioning has been extensively and accurately modeled by computer simulation ([4], for review). Barto and Sutton [5] comment in particular on the anticipatory nature of the process.

Even though they arise from profoundly different points of view, i.e. "behaviorist" vs. "cognitivist", there are many similarities between the operant conditioning and "cognitive" approaches. A key issue that separates them is the role of overt reward as a driver for learning. Is reward necessary for learning, as would be suggested by the operant conditioning approach? Clearly not, as indicated by the *latent learning* procedure ([21] for a review of the animal literature, and, e.g., [25], for a simulation using the DEM), in which rats (for instance) may be demonstrated to learn mazes in the absence of any externally applied reward. It is not until some rewarding condition is introduced into the maze that the same rats are observed to act in an obviously purposive manner within the maze. This, and similar observations, would suggest that learning and the motivation to utilize what is learnt are generally separate. It may, of course, be the case that an animal is partially or highly pre-disposed to learn combinations that are, have been, or might be "rewarding" (Witkowski, [25] models such an effect using the DEM).

Saksida *et al.* [14] present a computer model of operant conditioning for robot behavior shaping tasks. The Associative Control Process (*ACP*) model ([2]) develops the two-factor theorem of Mowrer ([12]). The ACP model reproduces a variety of

animal learning results from both classical and operant conditioning. Schmajuk [16] presents a two-part model combining both classical and operant conditioning modules emulating escape and avoidance learning behavior. Mowrer's work, combining aspects of classical conditioning and operant conditioning (the "two factors"), provides the inspiration for the title of this paper.

Several anticipatory and predictive three-part models have recently appeared in the Animat literature. Stolzmann *et al.* [18] describe an Anticipatory Classifier System (ACS), Witkowski ([25], [26], [27]) describes the Dynamic Expectancy Model (DEM). Developed independently, both are overtly predictive three-part systems, with a number of significant parallels and differences.

4 The Anticipatory Framework

This section proposes a framework of the three fundamental kinds of connection between stimulus Signs and Action response, and five basic rules relating their behavioral and predictive activities. The purpose of this section is to show that each of the four apparently disparate learning theories introduced in the previous sections can be unified from a single anticipatory or predictive viewpoint, and so how they might each serve a purpose within the individual animat.

Henceforth, the term sign-stimulus or simply *Sign* will be used to refer to an identifiably distinct conjunction of sensory conditions, all of which must be individually present for the Sign as a whole to be deemed *active*. A Sign that is predicted is referred to as *sub-active*, a status distinct from full activation as there are circumstances where anticipated activations must be treated differently from actual activation. The component parts of a Sign may be sensitive to a broad or narrow range of phenomena, and the Sign is active whenever each component is detecting anywhere in its range. The range of these components may be altered marginally at any given time. In principle, a Sign may detect external phenomena (as from a sensor or perceptual system), the activity status of an Action or a variety of other, internal, conditions. The total set of Signs currently known to the animat will be indicated by the calligraphic capital letter $\mathcal{S}$, an individual Sign by the lower case letter s and the active sub-set of Signs by $\mathcal{S}^*$.

The term *Action* (used from now on in preference to the pejorative, but largely synonymous term "response") will refer to recognizable units of activity performed by the animat, taken from the set of actions available to the animat. The animat will have a fixed repertoire of such action patterns (which may be simple or complex). Any action being currently expressed (performed) is deemed *active*. Actions may be overt, causing physical change to the animat's effector system or covert, specifically changing the status of a Sign's valence level or forming a connection between other Signs and Actions. The total set of Actions available to the animat will be indicated by the letter $\mathcal{A}$, and individual Actions by a_i, the active sub-set of Actions by $\mathcal{A}^*$.

The generally neutral term *valence* (after Tolman [23]) will be adopted to indicate that a Sign has goal like properties, in that it may give the appearance of driving or motivating the animat to activity. In this framework any Sign may have valence, which is separate property from activation or sub-activation. Like sub-activation it

may be propagated to other Signs; but does not, in itself, give rise to overt behavior. It remains unclear how motivation is derived in the brain, although its observable effects are clear enough.

We will assume that the animat has memory, conventionally, of past occurrences, but also a temporary ordered memory of predicted future occurrences. The extent of this memory (in terms of what may be recalled and the time period over which it is defined) will limit what may be learned and predicted. First we recognize three types of connection:

1) SA-Connection: Signs can be connected to Actions, either innately or as a consequence of learning.
2) SS-Connection: Signs may predict other Signs, where a predictive link has been established.
3) SAS-Connection: Signs may be attached to an Action and a second Sign, as prediction.

Next consider the following five "rules of propagation", which define (a) when an Action becomes a candidate for activation, (b) when a prediction will be made, (c) when a Sign will become sub-activated, and (d) when a Sign will become valenced:

1. When the Sign in an SA-Connection is active or sub-active the associated action becomes a candidate for activation (expression).
2. When the stimulus-sign in an SS-Connection is active or sub-active the consequent Sign becomes sub-active. Where the prediction implies a time delay, a future "memory" may be made of the predicted activation.
3. Any Sign that predicts another Sign (either SS or SAS) that has valence, itself becomes immediately valenced.
4. An SAS-Connection where the antecedent Sign and Action are both active is itself active and predicts its Consequent Sign, taking into account any delay.
5. The Action in an SAS-Connection where the antecedent Sign is both active and has valence (because its consequent Sign does, by rule 3) becomes a candidate for expression.

Rule 1: This is the standard behaviorist Stimulus-Response model. It may be applied to SA connections both in the sense of an Unconditioned Reflex in the classical conditioning domain, and in the sense of an action pattern releaser/trigger for a more complex behavior module, for instance using a "winner takes all strategy". As many Signs may be active at any one time and are not assumed to be mutually exclusive. It will also be assumed (in the absence of data to the contrary) that several UR may be initiated concurrently.

As action patterns become more complex the activation strategy becomes more critical. It is largely assumed that such complex activities are mutually exclusive (even though several Signs may be active), such that the activated behavior patterns will be in a priority order. The description of this process as a simple S-R activity belies the potential, and typical, complexity of the behaviors than can be initiated. Bryson's [8] EDMUND model, for instance, extends Rosenblatt and Payton's [13] feed-forward network model with elements of parallel activation and hierarchical control structures in order to explain the range of phenomena noted in nature.

Rule 2: Describes a simple predictive step, the occurrence of one Sign leading to the expectation or anticipation that a second will follow within a specified period. This rule accounts for the observations of classical (and higher order) conditioning phenomena (section 2.2) when in conjunction with rule one. Note that rule one only expresses the expression criteria in conditional terms, that sub-activation (the result of the predictive connection) may (or may not) activate the SA-connection. Despite an assumption of *equivalence of associability* (i.e. that any two Signs may act as either predictor or predictee, [6], p. 67), it is clear that not all stimuli are equally amenable to act as the CS in conditioning experiments. Shettleworth [17] found (in the case of golden hamsters) that it was easy to associate certain UR behaviors, such as "digging", "scrabbling" and rising up on the hind legs with a food outcome, and almost impossible to condition others, such as washing or scratching. Shettleworth also noted that the behaviors that could be conditioned were in any case those that the animal tended to emit ordinarily in anticipation of feeding, where the ones that could not be conditioned were not.

By rule 2, sub-activation is defined as self-propagating; sub-activation of the antecedent will in turn sub-activate the consequent. This defines the mechanism for longer chains, as would be the case in, for example, higher order conditioning. In some examples of second order conditioning schedules (e.g. Rizley and Rescorla, cited in [6]) it is possible to extinguish the initial (directly predicting) CS, without affecting the second-order CS. This appears consistent with the notion of propagating sub-activation, rather that full activation, which would indeed sever the chain.

A question remains as to the degree to which sub-activation should propagate in this manner. Given the reported difficulties of sustaining higher order conditioning schedules, it would seem plausible to suggest that propagated sub-activation in this sense will typically be a highly attenuating process in most instances. By treating sub-activation also as an anticipatory mechanism (in the Shettleworth [17] sense), that is, priming the animat for other activities, it would seem equally reasonable that the consequences of this predictive effect should remain localized. Without this restriction too many Signs would become sensitized and the effect would be diluted.

Rule 3: This rule describes the reverse effect of propagating valence (back) across a predictive link, from predictee to predictor. We may take the derived reinforcer as an exemplar of this process. Some Signs clearly have innate connection to the source of valence. That is, their occurrence predicts or is associated with a change in the state of the valence source. For a hungry dog, it seems that the taste or smell of meat has just such an effect. This is apparently innate and does not need to be established. By rule 3, the derived reinforcer, otherwise neutral, gains its valence by predicting that smell or taste. Clearly, the prediction link persists after the conditions that lead to its formation are lost.

Rule 3 applies to both SS and SAS type connections, as they are both overtly predictive forms. However they have different properties and should therefore propagate differently. The SAS connection, like a conventional problem-solving operator, implies action by the animat to move across the link. In this form the animat actively initiates the transitions. It is assumed that valence will propagate well across these links, capable of forming long chains of outcome predictions (section 2.3). Applying this rule rigorously, we note, however, that the propagation takes the form

of a graph between Signs linked by predictions. The sequence of actions it actually generates, on the other hand, will indeed appear as a linear sequence.

In the SS form the animat must essentially wait and see if the transition occurs. While useful for some schedules ("wait for the bell"), to rely on long chains of such connections would lead to effective behavioral paralysis. It is therefore assumed that valence, as with prediction, propagates poorly across SS connections.

Rule 4: Defines the conditions under which an SAS connection makes its prediction. Note that the prediction is made (and any sub-activations instigated) regardless of how the action was initiated.

Rule 5: Defines the conditions under which the action in an SAS connection itself becomes a candidate for activation. When an antecedent Sign is both active and has valence, it is at a point of intersection in the valence graph forming a plausible "chain of actions" to a source of valence (acting as a goal) from the animat's current situation. The Dynamic Expectancy Model takes into account the total (estimated) effort between each Sign and sources of valence by combining the effort that must be expended at each step with the strength of the prediction across the connection. Consequently the model defers action choice until the graph of Sign connections is completely evaluated, so as to be sure of selecting the action at the start of the most advantageous chain.

5 The DEM Postulates

The Dynamic Expectancy Model defines an animat controller based on the principles of the anticipatory approach described. The five "rules" discussed in section 4 serve to establish a broad framework. This section adds operational detail to those principles as a step to the computer program implementation of the model (SRS/E) in the form of a larger number of "postulates". Where the five rules emphasize the generation of predictions and the activation of behaviors from existing anticipatory connections, the postulates extend the discussion by considering how SS and SAS connections may be formed and maintained.

The anticipatory connections, SS and SAS, constitute a form of overtly predictive hypotheses. In the Dynamic Expectancy Model they will be referred to as *μ-hypotheses* (spoken "micro-hypotheses"). These are encapsulations of the two predictive, and so capable of corroboration "by experiment", forms (SS and SAS). Applications of these forms, where they make their prediction, will be considered as a form of experiment, or *μ-experiments* ("micro-experiments"). Each activation acts as a test to determine their overall effectiveness in representing the animat and its environment. The construction and corroboration of low-level observation based μ-hypotheses would appear a useful pre-cursor to the independent development of any systematic theoretical model, whose structure is not wholly or primarily dependent on an *originator* (the individual or process responsible for the creation of the animat and its ethogram).

5.1 The Hypothesis Postulates

Definition H0: The *μ-hypothesis*. Each of the forms SS and SAS shall be considered as μ-hypotheses, as each type is capable of forming a prediction and so is inherently "testable". Call the set of all μ-hypotheses $\mathcal{H}$, with h indicating an individual μ-hypothesis. A μ-hypothesis is composed of Signs (s) and Actions (a) from the respective Sign (S) and Action (A) lists. So:

SS: h_{SS}: $s' \overset{0}{\leftrightarrows}_{t\pm\tau} s''$ (eqn. 1)

SAS: h_{SAS}: $a \wedge s' \overset{0}{\leftrightarrows}_{t\pm\tau} s''$ (eqn. 2)

Each records a possible transition between two conditions that may be sensed by the animat (signs s' and s''). In an SAS connection s' must be concurrent ($\wedge$) with an action a. The double arrow ($\leftrightarrows$) now jointly indicates the left to right prediction (rules 2 and 4), of the consequent, and the instantaneous (rule 3) reverse transfer of valence.

Postulate H1: μ-Experimentation. μ-Experimentation is the mechanism by which predictive self-testability is conducted. μ-Experimentation is a two-part process. (1) making a prediction based on matching a μ-hypothesis' antecedent conditions to current activations, and (2) comparing those predictions, *a posteriori*, with the actual activations that hold true at the time stipulated by the prediction.

Postulate H2: Prediction. Prediction (implementing rules 2 and 4) records the predicted sign in prediction memory whenever a μ-hypothesis is active. Denoting the total set of active predictions made by the animat and currently awaiting confirmation with the letter $\mathcal{P}$, with p indicating an individual prediction. So:

SS: if ($h.s' \in S^*$) then $h.p^{+t} \leftarrow s''$ (eqn. 3)

SAS: if ($h.s' \in S^*$ & $h.a \in A^*$) then $h.p^{+t} \leftarrow s''$ (eqn. 4)

This memory is a property of the predicting μ-hypothesis, not of the sign predicted, as one Sign may be independently predicted by several μ-hypotheses.

In the SRS/E implementation, prediction memories are implemented as shift register like *traces*, the prediction being placed into the register $+t$ units ahead. The register moves one place backwards towards "the current time" with each execution cycle. The dot notation indicates that p is an attribute of h. Thus $h.p^{+t}$ indicates a prediction due at time t, made by the hypothesis h. A different implementation might record individually time stamped predictions, and so have an arbitrary time horizon.

Postulate H3: Corroboration. To match these recorded predictions against immediate sensations at the time the predictions fall due. If a μ-experiment is to be valid it must encapsulate all of the pre-conditions under which it will be judged. The antecedent components in a SS or SAS connection serve exactly as the definition of those pre-conditions. The quality of each μ-hypothesis is determined solely by its ability to accurately predict its consequent sign. This record of the animat's ability is encoded in the *corroboration measure* (C_m).

One might suppose that the corroboration measure is properly defined as the simple ratio of the total number of predictions made by the μ-hypothesis to the number of correct predictions made. This is equivalent to the probability (P_m), thus:

$P_m = p(\gamma'' |^t (\gamma' \wedge \alpha))$ (SAS form) (eqn. 5)

The use of the "*t*" symbol here acts as a reminder of the temporal relationship that exists between the predicted outcome and context. However, this measure is highly sensitive to sample size, if a μ-hypothesis were to change from being valid to invalid (the world changed) a long established μ-hypothesis would react slowly.

In practice, a confidence measure related to probability is adopted. Each successful prediction reinforces confidence in a μ-hypothesis. Conversely every unsuccessful prediction extinguishes confidence in that μ-hypothesis. The contributions of past predictions are discounted as further predictions are made and μ-hypotheses remain largely insensitive to their age and experience. The corroboration measure (C_m) is increased by the quantity:

$\Delta C_m = \alpha(1 - C_m)$ if $\eta : p^0 \in S^*$ (eqn. 6)

following each instance of a successful prediction of an active Sign, and

$\Delta C_m = - \beta(C_m)$ if $\eta : p^0 \notin S^*$ (eqn. 7)

following an unsuccessful prediction. C_m is updated following the widely used delta rule form. Under constant conditions these relationships give rise to the widely observed "negatively accelerating" form of the learning curve. Two proper fractions, the *reinforcement rate* (α) and the *extinction rate* (β) respectively define a "learning rate" for successful and unsuccessful prediction situations. They control the rate at which the influence of past predictions will be discounted. The C_m value of a μ-hypothesis that makes persistently successful predictions tends to 1.0, the C_m value of a μ-hypothesis that persistently makes unsuccessful predictions tends to 0.0. The positive reinforcement rate need not be equal to the negative extinction rate.

DEM μ-hypotheses are not derived from explicit theories, but are instead created from examples and may be thought of as "competing" to attain higher confidence measures and so be incorporated into goal-directed valence sequences and therefore influence the overt behavior of the animat.

Postulate H4: Learning by Creation. μ-Hypotheses may, of course, be innate to the animat, part of the ethogram definition. The prediction and corroboration mechanism will effectively tune them to the animat's actual circumstances. This both pre-disposes the animat to useful and (presumably) appropriate behavior patterns, and allows innate and learned behaviors to be integrated. However, to be a fully-fledged learning entity, the model must define a "Learning by Creation" method by which the animat extends the set of μ-hypotheses. This learning proceeds in two parts, (1) detecting circumstances where a new μ-hypothesis is required, and (2) the actions required to construct the double (SS) or triple (SAS) connection.

μ-Hypotheses exist to predict future occurrences of Signs; it is therefore reasonable to suppose that new μ-hypotheses should be created under two specific circumstances. Potentially, every sign should have at least one μ-hypothesis capable of predicting it, and ideally the Sign would be correctly predicted for every occurrence. Novel signs (ones not previously recognized by the system) can appear in the system as a result of the differentiation process (H5, below) where new, distinct Signs are formulated - **postulate H4-1**, *novel event*. In the second creation circumstance, known signs are detected without a corresponding prediction, **postulate H4-2**, *unexpected event*.

Novel and unexpected Signs are recognized within the SRS/E system by detecting the condition:

$\mathcal{s} \in \mathcal{S}^* \,\&\, \mathcal{s} \notin \mathcal{P}^0$, that is, the Sign $\mathcal{s}$ is active, but was not predicted to be so at the current time.

In either case a new μ-hypothesis may be created. The consequence Sign ($\mathcal{s}'$) for this new μ-hypothesis will be the novel or unexpected Sign. The context and action drawn from the set of recent Signs (and Actions for an SAS connection) recorded by the system in the memories associated with individual Signs and Actions (modeled in SRS/E as the shift register like "traces"). The new μ-hypothesis may then be constructed by incorporated the remembered components into the antecedent and shifting the predicted time by an amount equivalent to the depth in the memory trace of the antecedent item(s).

Postulate H5: Refinement. Refinement is the mechanism by which the animat may differentiate or generalize its existing set of μ-hypotheses. Differentiation adds extra conditions to the context of an existing μ-hypothesis, reducing the range of circumstances under which that μ-hypothesis will be applicable. Generalization removes or relaxes existing conditions to the context, increasing the range of circumstances. Differentiation may be appropriate to enhance μ-hypotheses that have stabilized, or stagnated, at some intermediate corroborative measure value. μ-Hypotheses should not be subject to differentiation until they have reached an appropriate level of testing (their "maturity", or extent of corroboration). Maturity is a measure of the degree of corroboration of a μ-hypothesis. It is otherwise independent of the age of a μ-hypothesis. It is expected that the refinement process will create new, separate μ-hypotheses that are derived from the existing ones. Both old and new μ-hypotheses are retained and may then "compete" to determine which offers the best predictive ability. In the specific implementation SRS/E, creation (H4) is heavily biased to formulating over-generalized μ-hypotheses, so differentiation is the primary refinement method. Anticipatory Classifier Systems (ACS), due to their design, tend to emphasize generalization [18] as the primary refinement mechanism.

Postulate H6: Forgetting. Forgetting is the mechanism by which the animat may discard μ-hypotheses found ineffective from the set of μ-hypotheses held, or when the system needs to recover resources. A μ-hypothesis might be deleted when it can be determined that it makes no significant contribution to the abilities of the animat. This point can be difficult to ascertain. Evidence from animal learning studies indicates that learned behaviors may be retained even after considerable periods of extinction. Experimental evidence drawn from the implementation of the Dynamic Expectancy Model points to the value of not prematurely deleting μ-hypothesis, even though their corroborative measures fall to very low levels [27]. Where a Sign is predicted by many μ-hypotheses there may be good cause to remove the least effective. It is presumed that the last remaining μ-hypothesis relating to a specific consequent Sign will not be removed, on the basis that some predictive ability, however poor, is better than none at all. As no record is retained of the forgotten μ-hypothesis, a new μ-hypothesis created later may be the same as one previously removed (by H4-2).

5.2 The Valence Postulates

Definition G0: Goals. A goal establishes a valence condition within the animat causing the animat to select behaviors appropriate to the achievement or *satisfaction* of that goal. Goals (denoted by the letters $\mathcal{G}/\mathcal{g}$) are a special condition of a Sign; goals are therefore always drawn from the set of available Signs.

Postulate G1: Goal Valence. From time to time the animat may assert any of the Signs available as a goal. Any Sign asserted to act as a goal in this way is termed as having valence (or be valenced). None, one or many Signs may be valenced at any one time.

Postulate G2: Goal Priority. Each valenced goal is assigned a positive, non-zero priority. This priority value indicates the relative importance to the animat of achieving this particular goal, in the prevailing context of other behaviors and goals. Goal priority is determined within the innate behavioral component of the ethogram. In the current SRS/E implementation only one goal is pursued at any time - the *top-goal*, the goal with the highest priority.

Postulate G3: Valenced Behavior. Whenever a goal is valenced, SRS/E will, by rule 4, propagate valence across existing μ-hypotheses to establish a graph of valenced connections within the system. In the SRS/E implementation each SRS connection will impose a *cost effort* estimate, C_e, proportional to the effort of performing the action and inversely to the current C_m value for the link:

$$C_e \leftarrow (\text{action\_cost}(\alpha) / C_m) \qquad \text{(eqn. 8)}$$

This effort accumulates across the graph, so that each antecedent Sign in the network defines the beginning of a path or chain (of actions) that represents the "best estimate" for the animat forward to the top-goal. This graph is referred to as the *Dynamic Policy Map* (DPM), as it defines a both preference ranking for activation for every Sign reached by rule 3 and indicates which of the actions associated by μ-hypotheses with the Sign should be activated. The DPM is recalculated frequently as goal priorities and confidence measures change due to corroboration, and as μ-hypotheses are added and removed from the system. In the SS connection case, it is convenient to consider a "dummy" action. By assigning it a high (notional) action_cost, propagation across these links is disadvantaged.

Postulate G4: Valenced Action Selection. When a DPM exists the system will apply rule 5 to activate a μ-hypothesis and so select an action. SRS/E selects the μ-hypotheses with the lowest overall cost estimate to the top-goal where several nodes compete for activation under rule 5.

Postulate G5: Goal Satisfaction. A valenced goal is deemed "satisfied" once the conditions defined by the goal are encountered, when the sign that defines the goal becomes activate. The priority of a satisfied goal is reduced to zero and it ceases to be a source of valence. Where goal-seeking behavior is to take the form of sustained maintenance of a goal state, the goal selection process must maintain the valence of the goal Sign following each satisfaction event.

Postulate G6: Goal Extinction. In a situation where all possible paths to a goal are unavailable, continued attempts to satisfy that goal will eventually become a threat to the continued survival of the animat, by blocking out other behaviors and

needlessly consuming resources. Such a goal must be forcibly abandoned. This is the *goal extinction point*. Witkowski [27] has modeled goal extinction using the DEM, arguing that is it substantially different from a simple reversal of the development of corroboration and from extinction in classical conditioning.

5.3 The Behavior Postulates

Definition B0: Behaviors. Behaviors (indicated by the letter $\mathcal{B}$) are non-learned activities inherent within the system. Behaviors are explicitly Stimulus-Response (SA) connections and are activated according to the tenets of rule 1. They are defined prior to parturition as part of the ethogram. There is no limit to the complexity (or simplicity) of innate behavior. An animat might be solely dependent on innate behaviors, with no learning component.

Postulate B1: Behavior Priority. Each behavior within the animat is assigned a priority relative to all the other behaviors. This priority is defined by the ethogram. The action associated with the behavior of highest priority is selected for expression.

Postulate B2: Primary Behaviors. Primary behaviors define the vocabulary of behavior patterns available to the animat at parturition. These behaviors provide a repertoire of activities enabling the animat to survive in its environment until learning processes may provide more effective behaviors.

Postulate B3: Goal Setting Behaviors. The ethogram defines the conditions under which the animat will convert to goal seeking behavior. Once a goal is set the animat is obliged to pursue that goal while there is no primary behavior of higher priority. Where no behavior can be selected from the DPM, the animat selects the primary behavior of highest priority that is currently active. Behavior selection from the DPM resumes once there is any match between the set of active signs and the current DPM.

Interruption of goal directed behavior by a higher priority innate behavior turns the animat away from pursuing its current top priority goal. For instance, goal directed food-seeking behavior should be interrupted by high priority predator avoidance activity. Once the threat has passed the goal directed behavior resumes, although the animat's perceived "place" in the DPM will have shifted as a result of the intervening behavior. The structure and corroboration of the DPM may have changed, and it must be re-evaluated as behavior reverts to the goal directed form. Where goal seeking takes the form of a sustained maintenance of the selected goal state, the selection process must re-valence the required goal each time it is satisfied.

Postulate B4: Default (exploratory) Behaviors. Default Behaviors provide a set of behaviors to be pursued by the animat whenever neither a primary nor a goal setting behavior is applicable. Typically these default behaviors will take the form of exploratory actions. Exploratory actions may be either random (*trial and error*), or represent a specific exploration strategy. Selection of this strategy will impact the rate and order in which the μ-hypothesis creation processes occur (H4). Default behaviors have a priority lower than any of the primary (B2) or goal setting (B3) behaviors. The provision of some default behaviors is mandatory within the ethogram.

6 The SRS/E Program Architecture

Figure one illustrates the flow of control within the SRS/E program architecture and the interaction between parts. The flow of control forms a non-terminating loop incorporating each of the eight steps identified in the figure. The computational effort of each cycle is relatively light, each activity being initiated opportunistically according to the prevailing circumstances. It is the cumulative effect over many cycles that gives rise, over time, to a refined set of corroborated μ-hypotheses.

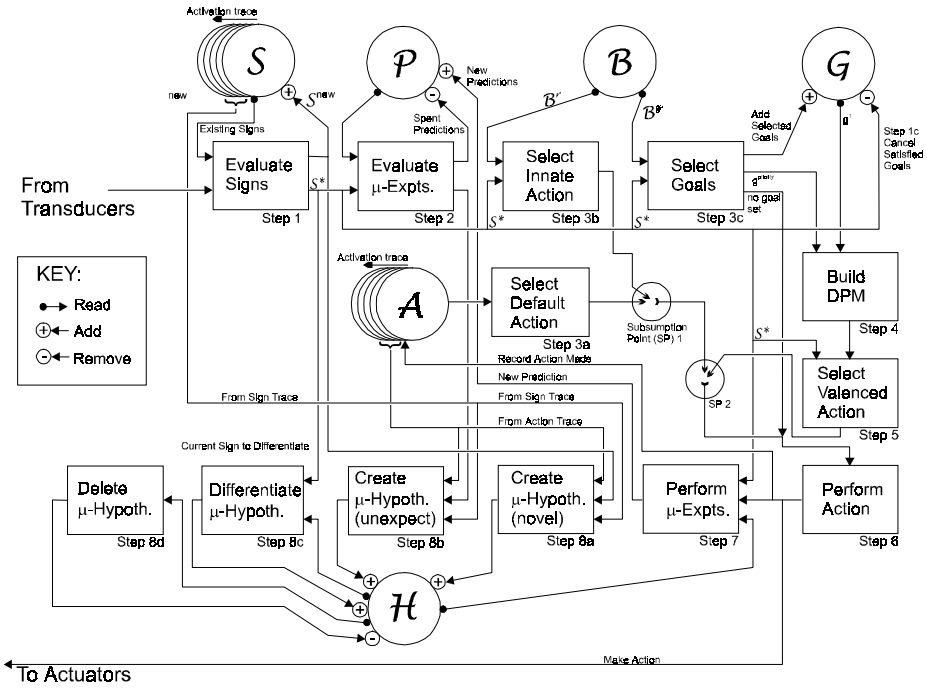

Figure One: The SRS/E Evaluation Cycle

Step 1 evaluates every sign to create the Sign activation list S^* using the current status of the animat's transducers. Step 2 compares past predictions falling due at the current time with the current activations and updates the corroboration measure of the μ-hypotheses responsible for the predictions tested (postulate H3). Step 3a selects a default (exploratory, B4) behavior. If an innate behavior is activated (postulate B2, step 3b) this will override the default behavior on the basis of priority (B1) at the *subsumption point* (SP1). Step 3c determines the valence status of Signs, and updates the Goal List (G), assigning each goal a priority (G2) on the basis of the defined goal setting behaviors (B3). Where at least one goal has valence (G1) step 4 is initiated and a Dynamic Policy Map constructed (G3). Step 5 applies rule 5 to find an intersection between S^* and μ-hypotheses valenced by step 4 (postulate G4). The highest priority action is passed (via subsumption point SP2) to step 6, which causes the animat's actuators to perform that action. Once an Action has been selected every μ-hypothesis

in $\mathcal{H}$ can be evaluated (postulates H1/H2) to determine the predictions to be made, which will be evaluated by step 2 in future cycles. Steps 8a, 8b, 8c and 8d implement postulates H4-1, H4-2, H5 and H6 respectively. The loop starts again at step 1.

7 An Illustrative Experimental Procedure

Figure two shows key stages from a single experimental run using the SRS/E program. The example illustrates a number of points arising from the use of the anticipatory learning approach described in this paper. First, demonstrating the use of anticipatory learning techniques to create new μ-hypotheses and corroborate them in the absence of explicit reward. Second, showing the effects of motivation on the behavior of the system and third, the effects of failed predictions in causing substantial changes to overt behavior during valenced behavior.

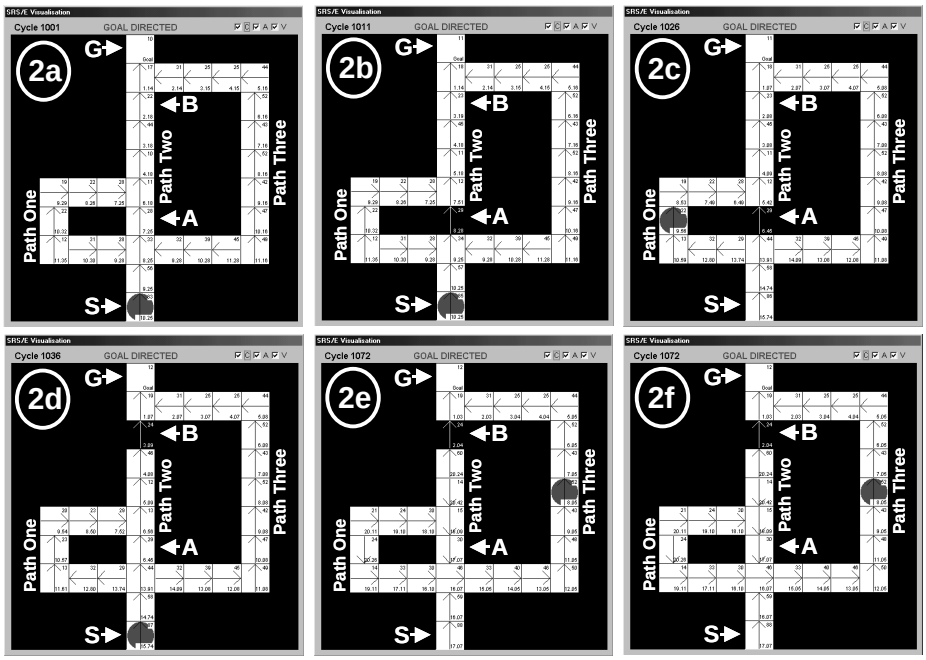

Figure two: Key Stages in the Experimental Procedure

The set-up represents a single animat (shown as the circular object) in a maze like environment where white squares represent traversable paths and black ones are blocked and cannot be entered. The animat may make one of four actions, moving "north" (up), "south", "east" or "west", taking it into an adjacent location. It does not move if the destination square is blocked or the edge of the environment is encountered. In this simulation the animat only senses a block when it attempts the move into it. The animat may directly and uniquely sense, as a Sign, the identity of the location it is currently occupying. The arrows represent the preferred action in the

event that the Sign representing a given location is encountered. The number in the top right corner the total number of times the location has been visited, that in the bottom corner the sum of cost effort estimates (C_e, eqn. 8) along the preferred path to the source of valence. The experimental procedure applied in the following stages ($\alpha = 0.5$, $\beta = 0.2$ throughout):

1) The animat is allowed to explore the maze for 1000 action cycles, but with no source of valence. Cycles 1 – 1000.
2) The animat is returned to the start 'S' and the Sign representing location 'G' is assigned valence by the experimenter. The animat is allowed to run to 'G'. In animal experiments, valence would be attached to the goal location by placing a "reward" (food, for example) there, acting as a secondary or derived reinforcer (section 2.3). Cycles 1001-1010.
3) A block is introduced at location 'A', the animat returned to 'S' and 'G' valenced. The animat is allowed to run to 'G'. Cycles 1011-1035.
4) The block at 'A' is removed and a block is now introduced at location 'B'. The animat is again returned to 'S' and 'G' valenced. The animat is allowed to run to 'G'. Cycles 1036-1080.
5) Stage 4 is repeated, 'B' remains blocked. Cycles 1081-1098.

During stage one of the experiment the animat uses the learning by creation (H4-1, H4-2) methods to formulate many μ-hypotheses relating to the environment and subsequently corroborates them using the methods described in postulates H1, H2 and H3. As there is no motivation or goal setting active during the first stage, all learning that takes place is "latent" (learning in the absence of explicit motivation or reward). The corroborative learning processes described by postulates H1, H2 and H3 depend only on the ability to make a prediction and so anticipate a specific outcome. A successful prediction is taken as its own reward. The system will also learn the maze if motivation is present, but, apparently paradoxically, learning may be less successful than under the unmotivated conditions ([27] for further discussion). This phenomenon has also been observed in the direct animal experiments on latent learning.

In stage two, the behavior of the animat is controlled by the Dynamic Policy Map, constructed by the repeated application of rule 3 as the Sign indicating the goal location 'G' now has valence. On the first run, the animat goes straight to the goal location via path two (figure 2a shows the DPM at cycle 1001). This is as expected. Stage three illustrates the effects of failed predictions, due to the introduction of a block at location 'A'. Figure 2b shows the situation at the start of stage 2 (cycle1011), anticipating a path via 'A'. On reaching the blockage, the animat attempts the action (North) to traverse into the expected location using the μ-hypothesis predicting the move. This fails, causing the corroboration measure for this μ-hypothesis to fall (eqn. 7) on successive attempts. As this individual hypothesis weakens, path two becomes less viable than path one (one may observe the DPM changing with each failed prediction and action). Figure 2c shows the situation at cycle 1026, with the animat now traversing path one. The visualization shows the situation after the route via path one becomes preferred, and the animat's behavior has changed to follow path one to 'G'. Figure 2d shows the DPM at the start of stage four. Due to the repeated failure of the μ-hypothesis entering location 'B', the animat backtracks to 'G' via path three once the blocked hypothesis is sufficiently weakened (figure 2e, cycle 1072). Figure

2f shows the situation at the start of stage five (cycle 1081). Inspection reveals that the DPM indicates a path to 'G' via path three. This is confirmed by the animat's actual path.

On an historical note, this procedure is based on the "place learning" procedure employed by Tolman and Honzik (see [6], p. 337) to demonstrate "inferential expectation" or "insight" in rats. The key question was whether the rat would take path two (as would be the case with a pure reward based reinforcement learning strategy) - the animat having no "insight" that the block at 'B' would also block path two) - or path three (the animat has a "map" strategy) at stage five. However, even this result relies on the animat being given sufficient time to fully explore its environment. Using a random or chance exploration strategy, insufficient exploration time can lead to non-optimal path choice preferences. This level of exploration is easily confirmed with the visualization tool (of figure two), but not so obvious with live rats. Where insufficient exploration is permitted, incomplete DPMs are formed and may not reach to the current location. The animat performs exploratory actions until a location with valence is encountered, when goal-seeking behavior resumes.

8 Summary and Conclusions

This paper has developed a minimal set of five "rules" of prediction and propagation by which an animat may exploit anticipation as a model of intelligence. The five rules are used to place the important attributes of four major learning and behavioral schemes (the "four factors") into a single anticipatory context and to develop a unified approach to modeling them. These are then elaborated with a larger number of "postulates", which act as a bridge to a realizable model (DEM) and the specific implementation (SRS/E). The key strength is the encapsulation of anticipatory prediction into μ-hypotheses, self-contained and capable of corroboration without recourse to any outside agency. Such μ-hypotheses anticipate what might happen (SS) and predict what can be made to happen (SAS), and can be used by the animat to derive appropriate behaviors in its environment as the need arises.

References

1. Agre, P.E.: Computational Research on Interaction and Agency, Artificial Intelligence, **72** (1995) 1-52
2. Baird, L.C. and Klopf, A.H.: Extensions to the Associative Control Process (ACP) Network: Hierarchies and Provable Optimality, 2nd Int. Conf. on Simulation of Adaptive Behavior (SAB-2), (1993) 163-171
3. Balkenius, C.: Natural Intelligence in Artificial Creatures, Lund University Cognitive Studies 37 (1995)
4. Balkenius, C. and Morén J.: Computational Models of Classical Conditioning: A Comparative Study, 5th Int. Conf. on Simulation of Adaptive Behavior (SAB-5), (1998) 348-353
5. Barto, A.G. and Sutton, R.S.: Simulation of Anticipatory Responses in Classical Conditioning by a Neuron-like Adaptive Element", Behavioral Brain Research, **4** (1982) 221-235

6. Bower, G.H. and Hilgard, E.R.: Theories of Learning, Englewood Cliffs: Prentice Hall Inc., fifth edition (1981)
7. Brooks, R.A.: Intelligence Without Reason, MIT AI Laboratory, A.I. Memo No. 1293. (Prepared for Computers and Thought, IJCAI-91, pre-print (April 1991)
8. Bryson, J.: Hierarchy and Sequence vs. Full Parallelism in Action Selection, 6th Int. Conf. on Simulation of Adaptive Behavior (SAB-6), (2000) 147-156
9. Butz, M.V., Sigaud, O. and Gerard, P.: Internal Models and Anticipations in Adaptive Learning Systems, Adaptive Behavior in Anticipatory Learning Systems 2002 Workshop (ABiALS 2002) 23pp.
10. Catania, A.C.: The Operant Behaviorism of B.F. Skinner, in: Catania, A.C. and Harnad, S. (eds.) "The Selection of Behavior", Cambridge: Cambridge University Press (1988) 3-8
11. Maes, P.: Behavior-based Artificial Intelligence, 2nd Int. Conf. on Simulation of Adaptive Behavior (SAB-2), (1993) 2-10
12. Mowrer, O.H.: Two-factor Learning Theory Reconsidered, with Special Reference to Secondary Reinforcement and the Concept of Habit, Psychological Review, **63**, (1956) 114-128
13. Rosenblatt, J.K. and Payton, D.W.: A Fine-Grained Alternative to the Subsumption Architecture for Mobile Robot Control, IEEE/INNS Int. Joint Conf. on Neural Networks, Vol. II. (1989) 317-323
14. Saksida, L.M., Raymond, S.M. and Touretzky, D.S.: Shaping Robot Behavior Using Principles from Instrumental Conditioning, Robotics and Autonomous Systems, **22**-3/4, (1997) 231-249
15. Schaffer, S.: Babbage's Intelligence: Calculating Engines and the Factory System, hosted at http://cci.wmin.ac.uk/schaffer/schaffer01.html (1998)
16. Schmajuk, N.A. Behavioral Dynamics of Escape and Avoidance: A Neural Network Approach, 3rd Int. Conf. on Simulation of Adaptive Behavior (SAB-3), (1994) 118-127
17. Shettleworth, S.J.: Reinforcement and the Organization of Behavior in Golden Hamsters: Hunger, Environment, and Food Reinforcement, Journal of Experimental Psychology: Animal Behavior Processes, **104**-1 (1975) 56-87
18. Stolzmann, W., Butz, M.V., Hoffmann, J. and Goldberg, D.E.: First Cognitive Capabilities in the Anticipatory Classifier System, 6th Int. Conf. on Simulation of Adaptive Behavior (SAB-6), (2000) 287-296
19. Sutton, R.S. and Barto, A.G.: Reinforcement Learning: An Introduction, Cambridge, MA: MIT Press (1998)
20. Tani, J. and Nolfi, S.: Learning to Perceive the World as Articulated: An Approach for Hierarchical Learning in Sensory-Motor Systems, 5th Int. Conf. on Simulation of Adaptive Behavior (SAB-5), (1998) 270-279
21. Thistlethwaite, D.: A Critical Review of Latent Learning and Related Experiments, Psychological Bulletin, **48**(2), (1951) 97-129
22. Thorndike, E.L.: Animal Intelligence: An Experimental Study of the Associative Processes in Animals, Psychol. Rev., Monogr. Suppl., **2**-8 (1898)
23. Tolman, E.C.: Purposive Behavior in Animals and Men, New York: The Century Co. (1932)
24. Tyrrell, T.: Computational Mechanisms for Action Selection, University of Edinburgh, Ph.D. thesis (1993)
25. Witkowski, M.: Dynamic Expectancy: An Approach to Behaviour Shaping Using a New Method of Reinforcement Learning, 6th Int. Symp. on Intelligent Robotic Systems, (1998) 73-81
26. Witkowski, M.: Integrating Unsupervised Learning, Motivation and Action Selection in an A-life Agent, 5th Euro. Conf. on Artificial Life (ECAL-99), (1999) 355-364
27. Witkowski, M. The Role of Behavioral Extinction in Animat Action Selection, proc. 6th Int. Conf. on Simulation of Adaptive Behavior (SAB-6), (2000) 177-186