Internal Models and Anticipations in Adaptive Learning Systems

Martin V. Butz[2,3], Olivier Sigaud[1], and Pierre Gérard[1]

[1] AnimatLab-LIP6, 8, rue du capitaine Scott, 75015 Paris France
{olivier.sigaud,pierre.gerard}@lip6.fr
[2] Department of Cognitive Psychology, University of Würzburg, Germany
butz@psychologie.uni-wuerzburg.de
[3] Illinois Genetic Algorithms Laboratory (IlliGAL),
University of Illinois at Urbana-Champaign, IL, USA

Abstract. The explicit investigation of anticipations in relation to adaptive behavior is a recent approach. This chapter first provides psychological background that motivates and inspires the study of anticipations in the adaptive behavior field. Next, a basic framework for the study of anticipations in adaptive behavior is suggested. Different anticipatory mechanisms are identified and characterized. First fundamental distinctions are drawn between implicit anticipatory behavior, payoff anticipatory behavior, sensory anticipatory behavior, and state anticipatory behavior. A case study allows further insights into the drawn distinctions. Many future research direction are suggested.

1 Introduction

The idea that *anticipations* influence and guide behavior has been increasingly appreciated over the last decades. Anticipations appear to play a major role in the coordination and realization of adaptive behavior. Various disciplines have explicitly recognized anticipations. For example, philosophy has been addressing our sense of reasoning, generalization, and association for a long time. More recently, experimental psychology confirmed the existence of anticipatory behavior processes in animals and humans over the last decades.

Although it might be true that over all constructible learning problems any learning mechanism will perform as good, or as bad, as any other one [71], the psychological findings suggest that in natural environments and natural problems learning and acting in an anticipatory fashion increases the chance of survival. Thus, in the quest of designing competent artificial animals, the so called *animats* [69], the incorporation of anticipatory mechanisms seems mandatory.

This book addresses two important questions of anticipatory behavior. On the one hand, we are interested in *how* anticipatory mechanisms can be incorporated in animats, that is, which structures and processes are necessary for anticipatory behavior. On the other hand, we are interested in *when* anticipatory mechanisms are actually helpful in animats, that is, which environmental preconditions favor anticipatory behavior.

To approach the *how* and *when*, it is necessary to distinguish first between different anticipatory mechanisms. With respect to the *how*, the question is *which* anticipatory mechanisms need which structure. With respect to the *when*, the question is *which* anticipatory mechanisms cause which learning and behavioral biases. In this chapter, we draw a first distinction between (1) *implicit anticipatory* mechanisms in which no actual predictions are made but the behavioral structure is constructed in an anticipatory fashion, (2) *payoff anticipatory* mechanisms in which the influence of future predictions on behavior is restricted to payoff predictions, (3) *sensory anticipatory* mechanisms in which future predictions influence sensory (pre-)processing, and (4) *state anticipatory* mechanisms in which predictions about future states directly influence current behavioral decision making. The distinctions are introduced and discussed within the general framework of *partially observable Markov decision processes* (POMDPs) and a general animat framework based on the POMDP structure.

The remainder of this chapter is structured as follows. First, psychology's knowledge about anticipations is sketched out. Next, we identify and classify different anticipatory mechanisms in the field of adaptive behavior. A non-exhaustive case study provides further insights into the different mechanisms as well as gives useful background for possible extensions. The conclusions outline many diverse future research directions tied to the study of anticipatory behavior in adaptive learning systems.

2 Background from Psychological Research

In order to motivate the usage of anticipations in adaptive behavior research, this section provides background from cognitive psychology. Starting from the behaviorist movement, we show how the notion of anticipation and its diverse impact on behavior was recognized in psychology research. While behaviorism gave rise to successful experimental psychology it somewhat ignored, and often even denied, anticipatory behavior influences. However, the experimental approach itself eventually revealed inevitable anticipatory influences on behavior. Recent neuron imaging techniques and single-cell recordings provide further proof of anticipatory cognitive processes.

2.1 Behaviorist Approach

Early suggestions of anticipations in behavior date back to Herbart [21]. He proposed that the "feeling" of a certain behavioral act actually triggers the execution of this act once the outcome is desired later.

The early 20th century, though, was dominated by the behaviorist approach that viewed behavior as basically stimulus-response driven. Two of the predominant principles in the behaviorist world are *classical conditioning* and *operant conditioning*.

Pavlov first introduced classical conditioning [39]. Classical conditioning studies how animals learn associations between an unconditioned stimulus (US) and a conditioned stimulus (CS). In the "Pavlovian dog", for example, the unconditioned stimulus (meat powder) leads to salivation — an unconditioned reflex (UR). After several experiments in which the sound of a bell (a neutral stimulus NS) is closely followed by the presentation of the meat powder, the dog starts salivating when it hears the sound of the bell independent of the meat powder. Thus the bell becomes a conditioned stimulus (CS) triggering the response of salivation.

While in classical conditioning the conditioned stimulus may be associated with the unconditioned stimulus (US) *or* with the unconditioned reflex (UR), operant conditioning investigates the direct association of behavior with favorable (or unfavorable) outcomes. Thorndike [60] monitored how hungry cats learn to escape from a cage giving rise to his "law of effect". That is, actions that lead to desired effects will be, other things being equal, associated with the situation of occurrence. The strength of the association depends on the degree of satisfaction and/or discomfort. More elaborate experiments of operant conditioning were later pursued in the well known "Skinner box" [46].

Thus, classical conditioning permits the creation of new CS on the basis of US, and operant conditioning permits to chain successive behaviors conditioned on different stimuli. Note that the learning processes take place backwards. To learn a sequence of behaviors, it is necessary to first learn the contingencies at the end of the sequence. In addition, the consequences are only learned because they represent punishments or rewards. Nothing is learned in the absence of any type of reward or punishment.

While behaviorism allowed cognitive psychology to make significant progress due to its principled study of behavior phenomena, a persisting drawback of the approach is the complete ignorance to, or denial of, any sort of mental state. Skinner's and others' mistake was to disallow future predictions or expectations, described as intentions, purposes, aims, or goals, to influence behavior.

> No one is surprised to hear it said that a person carrying good news walks more rapidly because he feels jubilant, or acts carelessly because of his impetuosity, or holds stubbornly to a course of action through sheer force of will. Careless references to purpose are still to be found in both physics and biology, but good practice has no place for them; yet almost everyone attributes human behavior to intentions, purposes, aims, and goals. [47, p.6]

Although Skinner is correct that the unscientific reference to e.g. "purpose" might result in the obstruction of scientific progress in psychology, we will show that it is possible to formalize future representations and behavior dependent on future representations. First, however, we present psychological investigations that clearly show that representations of the future are influencing behavior.

2.2 Expectancy Model

First experimental evidence for anticipatory behavior mechanisms can be found in Tolman's work [61–63]. Tolman proposed that, additionally to conditioned learning, *latent learning* takes place in animals. In latent learning experiments animals show to have learned an environmental representation during an exploration phase once a distinct reinforcer is introduced in the successive test phase (e.g. [61, 58]).

In typical latent learning experiments animals (usually rats) are allowed to explore a particular environment (such as a maze) without the provision of particular reinforcement. After the provision of a distinctive reinforcer, the animals show that they have learned an internal representation of the structure of the environment (by e.g. running straight to the food position).

More technically, the rats must have learned some environmental map (i.e., a predictive model) during exploration. Next, a goal emerges, that is, a certain state in the environment is desired. Finally, without any further active exploration, the rats are able to exploit the learned model and consequently move directly towards the desired state.

The observation of latent learning led Tolman to propose that animals form *expectancies*,

> [...] a condition in the organism which is equivalent to what in ordinary parlance we call a 'belief', a readiness or disposition, to the effect that an instance of this sort of stimulus situation, if reacted to by an instance of that sort of response, will lead to an instance of that sort of further stimulus situation, or else, simply by itself be accompanied, or followed, by an instance of that sort of stimulus situation.[64, p.113]

Essentially, expectancies are formed predicting action effects as well as stimulus effects regardless of actual reinforcement. A whole set of such expectancies, then, gives rise to a *predictive environmental model* which can be exploited for anticipatory behavior.

2.3 More Recent Psychological Evidence

In cognitive psychology anticipations have been experimentally shown to influence behavior ranging from simple reaction time tasks to elaborate reasoning tasks [29, 48]. It becomes more and more obvious that anticipations influence actual behavior as well as memory mechanisms and attention [37]. Neuropsychology gained further insights about the role of anticipatory properties of the brain in attentional mechanisms and, conversely, highlighted the role of attentional mechanisms in e.g. the anticipation of objects [43]. This section investigates two key findings in psychology research to show the broad impact of anticipatory behavior mechanisms.

Predictive capabilities come into play on different levels and to different extensions. The very recent discovery of *mirror neurons* in neuroscience provides neurological evidence that at least "higher" animals, such as monkeys, form representations of their conspecifics [41, 15]. The findings show that there are neurons in monkeys that are active not only when performing a particular action, such as grasping an object, but also when watching another monkey or human performing the same action. This shows that predicting the action of other people is realized by the re-use of neuronal pathways that represent one's own actions. For now, it is unclear how the other agent's actions are linked to ones own action representation. Gallese [14] suggests that the link may be constituted by the embodiment of the intended goal, shared by the agent and the observer. Gallese [14] also argues that only due to mirror neurons it may be possible to become socially involved enabling understanding and prediction of other people's intentions by a *shared manifold* — the association of other peoples actions and feelings with ones own actions and feelings via mirror neurons. Arbib [1] proposed mirror neurons as a prerequisite for the evolution of language. He suggests that it may only be possible to comprehend other people's speech acts by simulating and predicting these acts with neurons identical to ones own speech acts.

In general, mirror neurons are strongly related to the *simulation theory* of mind reading which postulates that in simulating other person's minds ones own resources are used. Simulation and prediction of other people's mind states mediated by mirror systems in the brain causes anticipatory behavior due to resulting predispositions in the mind. Empathy, for example, can be seen as a special case of anticipatory behavior in which motivational and emotional resources become active due to predictions and simulation of other people's minds by the means of mirror systems [59].

Another clear benefit can be found in research on attention. Pashler [38] gives a great overview over the latest research knowledge on attention in humans. LaBerge [31] distinguishes between *selective* and *preparatory attention*. While he suggests that selective attention does not require any anticipatory mechanisms, preparatory attention does. Preparatory attention predicts the occurrence of a visual perception (spatial or object-oriented) and consequently biases the filtering mechanism. The prediction is done by the system's model of its environment and influences the state of the system by the means of the decision maker's actions that essentially manipulate attentional mechanisms in this case. Preparatory attention enables faster goal-directed processing but may also lead to *inattentional blindness* [34]. In inattentional blindness experiments it is revealed that attention can be directed spatially, temporally, and/or object-oriented. It is most strikingly shown in the famous "gorilla experiment" [44]. A tradeoff arises between faster processing and focusing capabilities due to preparatory, or anticipatory, attention and a possible loss of important information due to inattention. When the capability of faster goal-directed processing outweighs the possibility of blindness effects needs to be addressed in further detail.

The next section introduces a formal framework for the classification of anticipatory mechanisms in animats and proposes first important distinctions.

3 Anticipation in Adaptive Behavior

Adaptive behavior is interested in how so called *animats* (artificial animals) can intelligently interact and learn in an artificial environment [69]. Research in artificial intelligence moved away from the traditional predicate logic and planning approaches to intelligence without representation [7]. The main idea is that intelligent behavior can arise without any high-level cognition. Smart connections from sensors to actuators can cause diverse, seemingly intelligent, behaviors. A big part of intelligence becomes *embodied* in the animat. It is only useful in the environment the animat is *situated* in. Thus, a big part of intelligent behavior of the animat arises from the direct interaction of agent architecture and structure in the environment.

As suggested in the psychology literature outlined above, however, not all intelligent behavior can be accounted for by such mechanisms. Thus, hybrid behavioral architectures are necessary in which an embodied intelligent agent may be endowed with higher "cognitive" mechanisms including developmental mechanisms, learning, reasoning, or planning. The resulting animat does not only act intelligently in an environment but it is also able to adapt to changes in the environment, to handle unforeseen situations, or to become socially involved. Essentially, the agent is able to learn and draw inferences by the means of internal representations and mechanisms. Anticipatory mechanisms may be part of these processes.

The cognitive mechanisms employed in animats are broad and difficult to classify and compare. Some animats might apply direct reinforcement learning mechanisms, adapting behavior based on past experiences but choosing actions solely based on current sensory input. Others might be enhanced by making actual action decisions also dependent on past perceptions. Anticipatory behavior research is interested in those animats that base their action decisions also on future predictions. Behavior becomes anticipatory in that predictions and beliefs about the future influence current behavior.

In the remainder of this section we develop a framework for animat research allowing for a proper differentiation of various types of anticipatory behavioral mechanisms. For this purpose, first the environment is defined as a partially observable Markov decision process (POMDP). Next, a general animat framework is outlined that acts upon the POMDP. Finally, anticipatory mechanisms are distinguished within the framework.

3.1 Framework of Environment

Before looking at the structure of animats, it is necessary to provide a general definition of which environment the animat will face. States and possible sensations in states need to be defined, actions and resulting state transitions need to be provided, and finally, the goal or task of the animat needs to be specified. The POMDP framework provides a good means for a general definition of such environments.

We define a POMDP by the $< X, Y, U, T, O, R >$ tuple

- X, the state space of the environment;
- Y, the set of possible sensations in the environment;
- U, the set of possible actions in the environment;

- $T : X \times U \rightarrow \Pi(X)$ the state transition function, where $\Pi(X)$ is the set of all probability distributions over X;
- $O : X \rightarrow \Pi(Y)$ the observation function, where $\Pi(Y)$ is the set of all probability distributions over Y;
- $R : X \times U \times X \rightarrow \mathbb{R}^r$ the immediate payoff function, where r is the number of criteria;

A Markov decision process (MDP) is given when the Markov property holds: the effects of an action solely depend on current input. Thus, the POMDP defined above reduces to an MDP if each possible sensation in the current state uniquely identifies the current state. That is, each possible sensation in a state x (i.e., all $y \in Y$ for which $O(x)$ is greater than zero) is only possible in this state. If an observation does not uniquely identify the current state but rather provides an (implicit) probability distribution over possible states, the Markov property is violated and the environment turns into a *non-Markov problem.* In this case, optimal action choices do not necessarily depend only on current sensory input anymore but usually depend also on the history of perceptions, actions, and payoff.

3.2 Adaptive Agent Framework

Given the environmental properties, we sketch a general animat framework in this section. We define an animat by a 5-tuple $\mathcal{A} =< S, A, M^S, M^P, \Pi >$. This animat acts in the above defined POMDP environment.

At a certain time t, the animat perceives sensation $y(t) \in Y$ and reinforcement $P(t) \in \mathbb{R}$. The probability of perceiving $y(t)$ is determined by the probability vector $O(x(t))$ and similarly, the probability of $x(t)$ is determined by the probability vector $T(x(t-1), u(t-1))$ which depends on the previous environmental state and the executed action. The received reward depends on the executed action as well as the previous and current state, $P(t) = R(x(t-1), u(t-1), x(t))$.

Thus, in a behavioral act an animat $\mathcal{A}$ receives sensation $y(t)$ and reinforcement $P(t)$ and chooses to execute an action A. To be able to learn and reason about the environment, $\mathcal{A}$ has internal states denoted by S that can represent memory of previous interactions, current beliefs, motivations, intentions etc. Actions $A \subseteq U$ denote the action possibilities of the animat. For our purposes separated from the internal state, we define a state model M^S and a predictive model M^P. The state model M^S represents current environmental characteristics the agent believes in — an implicit probability distribution over all possible environmental states X. The predictive model M^P specifies how the state model changes, possibly dependent on actions. Thus, it describes an implicit and partially action-dependent probability distribution of future environmental states. Finally, Π denotes the behavioral policy of the animat, that is, how the animat decides on what to do, or which action to execute. The policy might depend on current sensory input, on predictions generated by the predictive model, on the state model, and on the internal state.

Learning can be incorporated in the animat by allowing the modification of the components over time. The change of its internal state could, for example, reflect the gathering of memory or the change of moods. The state model could be modified by generalizing over, for example, equally relevant sensory input. The predictive model could learn and adapt probabilities of possible state transitions as well as generalize over effects and conditions.

This rather informal agent framework suffices for our purposes of distinguishing between different classes of anticipatory behavior in animats.

3.3 Distinctions of Anticipatory Behavior

Within the animat framework above, we can infer that the predictive model M^P plays a major role in anticipatory animats. However, in the broader sense of anticipatory behavior also animats without such a model might be termed anticipatory in that their behavioral program is constructed in anticipation of possible environmental challenges. We term this first class of anticipations implicitly anticipatory. The other three classes utilize some kind of prediction to influence behavior. We distinguish between payoff anticipations, sensory anticipations, and state anticipations. All four types of anticipatory behavior are discussed in further detail below.

Implicitly Anticipatory Animats The first animat-type is the one in which no predictions whatsoever are made about the future that might influence the animat's behavioral decision making. Sensory input, possibly combined with internal state information, is directly mapped onto an action decision. The predictive model of the animat M^P is empty or does not influence behavioral decision making in any way. Moreover, there is no action comparison, estimation of action-benefit, or any other type of prediction that might influence the behavioral decision. However, implicit anticipations are included in the behavioral program of the animat. The basic structure of an implicit anticipatory mechanism is shown in Figure 1.

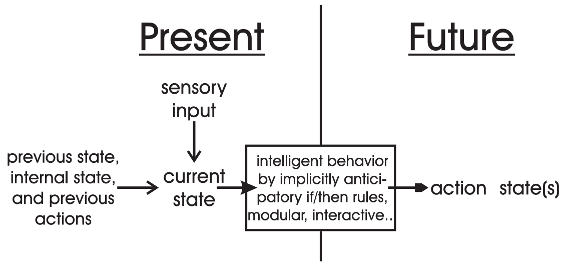

Fig. 1. Implicit anticipatory behavior does not rely on any explicit knowledge about possible future states. The behavior is anticipatory in that the behavioral architecture is predicted to be effective. For example, a genetic code is implicitly predicted (by evolution) to result in successful survival and reproduction.

In nature, even if a life-form behaves purely reactively, it has still implicit anticipatory information in its genetic code in that the behavioral programs in the code are (implicitly) anticipated to work in the offspring. Evolution is the implicit anticipatory learning mechanism that imprints implicit anticipations in the genes. Similarly, well-designed implicitly anticipatory animats, albeit without any prediction that might influence behavior, have implicit anticipatory information in the structure and interaction of algorithm, sensors, and actuators. The designer has included implicit anticipations of environmental challenges and behavioral consequences in the controller of the animat.

It is interesting to note that this rather broad understanding of the term "anticipation" basically classifies any form of life in this world as either implicitly anticipatory or more explicitly anticipatory. Moreover, any somewhat successful animat program can be classified as implicitly anticipatory since its programmed behavioral biases are successful in the addressed problems. Similarly, any meaningful learning mechanism works because it supposes that future experience will be somewhat similar to experience in the past and consequently biases its learning mechanisms on experience in the past. Thus, any meaningful learning and behavior is implicitly anticipatory in that it anticipates that past knowledge and experience will be useful in the future. It is necessary to understand the difference between such implicitly anticipatory animats and animats in which explicit future representations influence behavior.

Payoff Anticipations If an animat considers predictions of the possible payoff of different actions to decide on which action to execute, it may be termed payoff anticipatory. In these animats, predictions estimate the benefit of each possible action and bias action decision making accordingly. No state predictions influence action decision making. A payoff anticipatory mechanism is schematized in Figure 2.

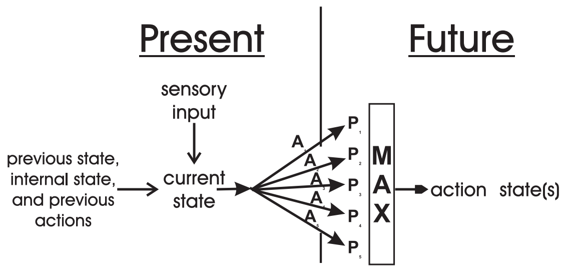

Fig. 2. Sensory anticipatory behavior influences sensory processing due to sensory predictions, expectations, or goal-dependent relevance measures.

A particular example for payoff anticipations is direct (or model-free) reinforcement learning (RL). Hereby, payoff is estimated with respect to the current behavioral strategy or in terms of possible actions. The evaluation of the estimate causes the alternation of behavior which again cause the alternation of the payoff estimates. It can be distinguished between on-policy RL algorithms, such as the SARSA algorithm [42, 52], and off-policy RL algorithms, such as Q-learning [65, 52] or recent learning classifier systems such as XCS [67].

Sensorial Anticipations While in payoff anticipations predictions are restricted to payoff, in sensory anticipations predictions are unrestricted. However, sensory anticipations do not influence the behavior of an animat directly but sensory processing is influenced. The prediction of future states and thus the prediction of future stimuli influences stimulus processing. To be able to form predictions, the animat must use a (not necessarily complete) predictive model M^P of its environment (see Section 3.2). Expected sensory input might be processed faster than unexpected input or unexpected input with certain properties (for example possible threat) might be reacted to faster. A sensory anticipatory mechanism is sketched in Figure 3.

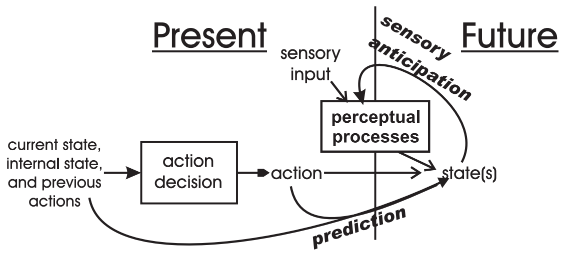

Fig. 3. Sensory anticipatory behavior influences, or predisposes, sensory processing due to future predictions, expectations, or intentions.

Sensory anticipations strongly relate to preparatory attention in psychology [31, 38] in which top-down processes such as task-related expectations influence sensory processing. Behavior is not directly influenced but sensory (pre-)processing is. In other words, sensory anticipatory behavior results in a predisposition of processing sensory input. For example, the agent may become more susceptible to specific sensory input and more ignorant to other sensory input. The biased sensory processing might then (indirectly) influence actual behavior. Also learning might be affected by such a bias as suggested in psychological studies on learning [22, 48].

State Anticipations Maybe the most interesting group of anticipations is the one in which animat behavior is influenced by explicit future state representations. As in sensory anticipations, a predictive model M^P must be available to the animat or it must be learned by the animat. In difference to sensory anticipations, however, state anticipations directly influence current behavioral decision making. Explicit anticipatory behavior is schematized in figure 4. The essential property is that prediction(s) about, or simply representations of, future state(s) influence actual action decision.

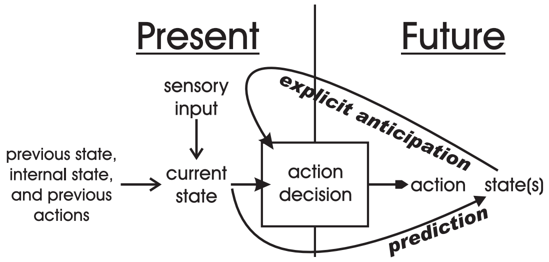

Fig. 4. Explicit anticipations influence actual action decision making due to future predictions, expectations, or intentions.

The simplest kind of explicit anticipatory animat would be an animat which is provided with an explicit predictive model of its environment. The model could be used directly to pursue actual goals by the means of explicit planning mechanisms such as diverse search methods or *dynamic programming* [5]. The most extreme cases of such high-level planning approaches can be found in early artificial intelligence work such as the general problem solver [36] or the STRIPS language [13]. Nowadays, somewhat related approaches try to focus on *local mechanisms* that extract only *relevant environmental information*.

In RL, for example, the dynamic programming idea was modified yielding indirect (or model-based) RL animats. These animats learn an explicit predictive model of the environment. Decisions are based on the predictions of all possible behavioral consequences and essentially the utility of the predicted results. Thus, explicit representations of future states determine behavior.

Further distinctions in state anticipatory animats are evident in the structure and completeness of the model representation, the learning and generalization mechanisms that may change the model over time, and the mechanisms that exploit the predictive model knowledge to adapt behavior. The structure of the predictive model can be represented by rules, by a probabilistic network, in the form of hierarchies and so forth. The model representation can be based on internal model states $M^S(t)$ or rather directly on current sensory input $y(t)$. State information in the sensory input can provide global state information or rather local state information dependent on the animat's current position in the environment. Learning and generalization mechanisms give rise to further crucial differences in the availability, the efficiency, and the utility of the predictive model. Finally, the bias of the behavioral component results in different anticipatory behavior mechanisms. For example, the number of steps that the animat can look into the future is a crucial measure as proposed in [45]. Moreover, anticipatory processes might only take place in the event of actual behavioral execution or the processes may be involved in adapting behavior offline. Proper distinctions between these different facets of state anticipatory behavior may be developed in future research.

With a proper definition of animats and four fundamental classes of anticipatory behavior in hand, we now provide a case study of typical existing anticipatory animats.

4 Payoff Anticipatory Animats

This section introduces several common payoff anticipatory animats. As defined above, these animats do not represent or learn a predictive model M^P of their environment but a knowledge base assigns values to actions based on which action decisions are made.

4.1 Model-Free Reinforcement Learning

The reinforcement learning framework [27, 52] considers adaptive agents involved in a sensory-motor loop acting upon a MDP as introduced above (extensions to POMDPs can be found for example in [9]). The task of the agents is to learn an optimal policy, i.e., how to act in every situation in order to maximize the cumulative reward over the long run.

In model-free RL, or *direct reinforcement learning*, the animat learns a behavioral policy without learning an explicit predictive model. The most common form of direct reinforcement learning is to learn utility values for all possible state-action combinations in the MDP. The most common approach in this respect is the Q-learning approach introduced in [65]. Q-learning has the additional advantage that it is policy independent. That is, as long as the behavioral policy assures that all possible state action transitions are visited infinitely often over the long run, Q-learning is guaranteed to generate an optimal policy.

Model-free RL agents are clearly payoff anticipatory animats. There is no explicit predictive model; however, the learned reinforcement values estimate action-payoff. Thus, although the animat does not explicitly learn a representation with which it knows the actual sensory consequences of an action, it can compare available action choices based on the payoff predictions and thus act payoff anticipatory.

Model-free RL in its purest form usually stores all possible state-action combinations in tabular form. Also, states are usually characterized by unique identifiers rather than by sensory inputs that allow the identification of states. This ungeneralized exhaustive state representation prevents RL to scale-up to larger problems. Several approaches exist that try to overcome the curse of dimensionality by function approximation techniques (cf. [52]), hierarchical approaches (cf. [54, 4]), or online generalization mechanisms. Approaches that generalize online over sensory inputs (for example in the form of a feature vector) are introduced in the following.

4.2 Learning Classifier Systems

Learning Classifier Systems (LCSs) have often been overlooked in the research area of RL due to the many interacting mechanisms in these systems. However, in their purest form, LCSs can be characterized as RL systems that generalize online over sensory input. This generalization mechanism leads to several additional problems especially with respect to a proper propagation of RL values over the whole state action space.

The first implementation of an LCS, called CS1, can be found in [25]. Holland's goal was to propose a model of a cognitive system that is able to learn using both reinforcement learning processes and genetic algorithms [23, 20]. The first systems, however, were rather complicated and lacked efficiency.

Reinforcement values in LCSs are stored in a set (the population) of condition-action rules (the classifiers). The conditions specify a subset of possible sensations in which the classifier is applicable thus giving rise to focusing mechanisms and attentional mechanisms often over-looked in RL. The learning mechanism of the population of classifiers and the classifier structure is usually accomplished by the means of a genetic algorithm (GA). Lanzi provides an insightful comparison between RL and learning classifier systems [33]. It appears from this perspective that a LCS is a rule-based reinforcement learning system endowed with the capability to generalize what it learns.

Thus, also LCSs can be classified as payoff-anticipatory animats. The generalization over the perceptions promises faster adaptation in dynamic environments. Moreover, the policy representation may be more compact especially in environments in which a lot of sensations are available but only a subset of the sensations is task relevant.

Recently, Wilson implemented several improvements in the LCS model. He modified the traditional Bucket Brigade algorithm [26] to resemble the Q-learning mechanism propagating Q-values over the population of classifiers [66, 67]. Moreover, Wilson drastically simplified the LCS model [66]. Then, he modified Holland's original strength-based criterion for learning — the more a rule receives reward (on average), the more fit it is [23, 24, 6] — by a new criterion relying on the accuracy of the reward prediction of each rule [67]. This last modification gave rise to the most commonly used LCS today, XCS.

5 Anticipations Based on Predictive Models

While the model-free reinforcement learning approach as well as LCSs do not have or use a predictive model representation, the agent architectures in this section all learn or have a predictive model M^P and use this model to yield anticipatory behavior. Due to the usage of an explicit predictive model of the environment, all systems can be classified as either sensory anticipatory or state anticipatory. Important differences of the systems are outlined below.

5.1 Model-based Reinforcement Learning

The dynamical architecture *Dyna* [53] learns a model of its environment in addition to reinforcement values (state values or Q-values). Several anticipatory mechanisms can be applied such as biasing the decision maker toward the exploration of unknown/unseen regions or applying internal reinforcement updates. Dyna is one of the first state anticipatory animat implementations. It usually forms an ungeneralized representation of its environment in tabular form but it is not necessarily restricted to such a representation. Interesting enhancements of Dyna have been undertaken optimizing the internal model-based RL process [35, 40] or adopting the mechanism to a tile coding approach [30]. The introduction of Dyna was kept very general so that many of the subsequent mechanisms can be characterized as Dyna mechanisms as well. Differences can be found in the learning mechanism of the predictive model, the sensory input provided, and the behavioral policy learning.

5.2 Schema Mechanism

An architecture similar to the Dyna architecture was published in [11]. The implemented *schema mechanism* is loosely based on Piaget's proposed developmental stages. The model in the schema mechanism is represented by rules. It is learned bottom-up by generating more specialized rules where necessary. Although no generalization mechanism applies, the resulting predictive model is somewhat more general than a tabular model. The decision maker is — among other criteria — biased on the exploitation of the model to achieve desired items in the environment. Similar to Dyna, the schema mechanism represents an explicit anticipatory agent. However, the decision maker, the model learner, and the predictive model representation M^P have a different structure.

5.3 Expectancy Model SRS/E

Witkowski [70] approaches the same problem from a cognitive perspective giving rise to his *expectancy model SRS/E*. Similar to Dyna, the learned model is not generalized but represented by a set of rules. Generalization mechanisms are suggested but not tested. SRS/E includes an additional sign list that stores all states encountered so far. In contrast to Dyna, reinforcement is not propagated online but is only propagated once a desired state is generated by a behavioral module. The propagation is accomplished using dynamic programming techniques applied to the learned predictive model and the sign list.

5.4 Anticipatory Learning Classifier Systems

Similar to the schema mechanism and SRS/E, anticipatory learning classifier systems (ALCSs) [50, 8, 19, 17] contain an explicit prediction component. The predictive model consists of a set of rules (classifiers) which are endowed with a so called "effect" part. The effect part predicts the next situation the agent will encounter if the action specified by the rules is executed. The second major characteristic of ALCSs is that they generalize over sensory input.

ACS An *anticipatory classifier system* (ACS) was developed by Stolzmann [49, 50] and was later extended to its current state of the art, ACS2 [8]. ACS2 learns a generalized model of its environment applying directed specialization as well as genetic generalization mechanisms. It has been experimentally shown that ACS2 reliably learns a complete, accurate, and compact predictive model of several typical MDP environments. Reinforcement is propagated directly inside the predictive model resulting in a possible model aliasing problem [8]. It was shown that ACS2 mimics the psychological results of latent learning experiments as well as outcome devaluation experiments mentioned above by implementing additional anticipatory mechanisms into the decision maker [50, 51, 8].

YACS *Yet Another Classifier System* (YACS) is another anticipatory learning classifier system that forms a similar generalized model applying directed specialization as well as generalization mechanisms [17, 18]. Similar to SRS/E, YACS keeps a list of all states encountered so far. Unlike SRS/E, reinforcement updates in the state list are done while interacting with the environment making use of the current predictive model. Thus, YACS is similar to SRS/E but it evolves a more generalized predictive model and updates the state list online.

MACS A more recent approach by [16] learns a different rule-based representation in which rules are learned separately for the prediction of each sensory attribute. Similar to YACS, MACS keeps a state list of all so far encountered states and updates reinforcement learning in those states. The different model representation is shown to allow further generalizations in maze problems.

5.5 Artificial Neural Network Models of Anticipation

Also Artificial Neural Networks (ANN) can be used to learn the controller of an agent. In accordance with the POMDP framework, the controller is provided with some inputs from the sensors of the agent and must send some outputs to the actuators of the agent. Learning to control the agent consists in learning to associate the good set of outputs to any set of inputs that the agent may experience.

The most common way to perform such learning with an ANN consists in using the back-propagation algorithm. This algorithm consists in computing for each set of inputs the errors on the outputs of the controller. With respect to the computed error, the weights of the connections in the network are modified so that the error will be smaller the next time the same inputs are encountered.

The main drawback of this algorithm is that one must be able to decide for any input what the correct output should be so as to compute an error. The learning agent must be provided with a supervisor which tells at each time step what the agent should have done. Back-propagation is a supervised learning method. The problem with such a method is that in most control problems, the correct behavior is not known in advance. As a consequence, it is difficult to build a supervisor.

The solution to this problem consists in relying on anticipation [55, 57]. If the role of an ANN is to predict what the next input will be rather than to provide an output, then the error signal is available: it consists in the difference between what the ANN predicted and what has actually happened. As a consequence, learning to predict thanks to a back-propagation algorithm is straight-forward.

Baluja's Attention Mechanism Baluja and Pomerleau provide an interesting anticipatory implementation of visual attention in the form of a neural network with one hidden layer [2, 3]. The mechanism is based on the ideas of visual attention modeling in [28]. The system is for example able to learn to follow a line by the means of the network. Performance of the net is improved by adding another output layer, connected to the hidden layer, which learns to predict successive sensory input. Since this output layer is not used to update the weights in the hidden layer, Baluja argues that consequently the predictive output layer can only learn task-relevant predictions. The predictions of the output layer are used to modify the successive input in that the strong differences between prediction and real input are decreased assuming strong differences to be task irrelevant noise. Baluja shows that the neural net is able to utilize this image flattening to improve performance and essentially ignore spurious line markings and other distracting noise. It is furthermore suggested that the architecture could also be used to detect unexpected sensations faster possibly usable for anomaly detection tasks.

Baluja's system is a payoff anticipatory system. The system learns a predictive model which is based on pre-processed information in the hidden units. The predictive model is action-independent. Sensory anticipations are realized in that the sensory input is modified according to the difference between predicted and actual input.

Tani's Recurrent Neural Networks Tani published a recurrent neural network (RNN) approach implementing model-based learning and planning in the network [55]. The system learns a predictive model using the sensory information of the next situation as the supervision. *Context units* are added that feed back the values of the current hidden units to additional input units. This recurrence allows a certain internal representation of time [12]. In order to use the emerging predictive model successfully, it is necessary that the RNN becomes situated in the environment — the RNN needs to identify its current situation in the environment by adjusting its recurrent inputs. Once the model is learned, a navigation phase is initiated in which the network is used to plan a path to a provided goal.

The most appealing result of this work is that the RNN is actually implemented in a real mobile robot. The implementation is shown to handle noisy, online discretized environments. Anticipatory behavior is implemented by a lookahead planning mechanism. The system is a state anticipatory system in which the predictive model is represented in a RNN. In contrast to the approaches above, the RNN also evolves an implicit state model M^S represented and updated by the recurrent neural network inputs. This is the reason why the network has to become situated before planning is applicable. Tani shows that predicting the next inputs correctly helps stabilizing the behavior of its agents and, more generally, that using anticipations results in a bi-polarization of the behavior into two extreme modes: a very stable mode when everything is predicted correctly, and a chaotic mode when the predictions get wrong.

In a further publication [56], Tani uses a constructivist approach in which several neural networks are combined. The approach implements an attentional mechanism that switches between wall following and object recognition. Similar to the winner-takes-all algorithm proposed in [28], Tani uses a winner-takes-all algorithm to implement a visual attention mechanism. The algorithm combines sensory information with model prediction, thus pre-processing sensory information due to predictions. The resulting categorical output influences the decision maker that controls robot movement. Thus, the constructed animat comprises sensory anticipatory mechanisms that influence attentional mechanisms similar to Baluja's visual attention mechanism but embedded in a bigger modular structure.

In [57], a first approach of a hierarchical structured neural network suitable as a predictive model is published. While the lower level in the hierarchy learns the basic sensory-motor flow, the higher level learns to predict the switching of the network in the lower level and thus a more higher level representation of the encountered environment. Anticipatory behavior was not shown within the system.

5.6 Anticipations in a Multi-Agent Problem

A first approach that combines low level reactive behavior with high-level deliberation can be found in [10]. The animats in this framework are endowed with a predictive model that predicts behavior of the other, similar animats. Although the system does not apply any learning methods, it is a first approach of state anticipations in a multi-agent environment. It is shown that by anticipating the behavior of the other agents, behavior can be optimized achieving cooperative behavior. Davidsson's agent is a simple anticipatory agent that uses the (restricted) predictive model of other agents to modify the otherwise reactive decision maker. Since the decision maker is influenced by the predictive model the agents can be classified as non-learning state-anticipatory animats.

6 Discussion

As can be seen in the above study of anticipatory systems, a lot of research is still needed to clearly understand the utility of anticipations. This section further discusses different aspects in anticipatory approaches.

6.1 Anticipating With or Without a Model

One main advantage of model building animats with respect to model-free ones is that their model endows them with a planning capability. Having an internal predictive model which specifies which action leads from what state to what other state permits the agent to plan its behavior “in its head”. But planning does not necessarily mean that the agent actually searches in its model a complete path from its current situation to its current goal. Indeed, that strategy suffers from a combinatorial explosion problem. It may rather mean that the agent updates the values of different state model states ($x \in M^S$) without having to actually move in its environment. This is essentially done in dynamic programming [5] and it is adapted to the RL framework in the Dyna architecture [53, 52]. The internal updates allow a faster convergence of the learning algorithms due to the general acceleration of value updates.

These ideas have been re-used in most anticipatory rule-based learning systems described above. Applying the same idea in the context of ANN, with the model being implemented in the weights of recurrent connections in the network, would consist in letting the weights of the recurrent connections evolve faster than the sensory-motor dynamics of the network. To our knowledge, though, this way to proceed has not been used in any anticipatory ANN animat, yet.

Pros and Cons of Anticipatory Learning Classifier Systems Having an explicit predictive part in the rules of ALCSs permits a more directed use of more information from the agent's experience to improve the rules with respect to classical LCSs. Supervised learning methods can be applied. Thus, there is a tendency in ALCSs to use heuristic search methods rather than blind genetic algorithms to improve the rules.

This use of heuristic search methods results then in a much faster convergence of anticipatory systems on problems where classical LCSs are quite slow, but it also results in more complicated systems, more difficult to program, and also in less general systems.

For example, XCS-like systems can be applied both to single-step problems such as Data Mining Problems [68] where the agent has to make only one decision independent from its previous decisions and to multi-step problems where the agent must run a sequence of actions to reach its goal [32]. In contrast, ALCSs are explicitly devoted to multi-step problems, since there must be a “next” situation after each action decision from the agent.

6.2 A Parallel Between Learning Thanks to Prediction in ANN and in ALCS

The second matter of discussion emerging from this overview is the parallel that can be made in the way ANN and rule-based systems combine predictions and learning to build and generalize a model of the problem.

We have seen that in Tani's system, the errors on predictions are back-propagated through the RNN so as to update the weights of the connections.

This learning process results in an improved ability to predict, thus in a better predictive model.

The learning algorithms in the presented ALCSs rely on the same idea. The prediction errors are represented by the fact that the predictions of a classifier are sometimes good and sometimes bad, in which case the classifier oscillates (or is called not reliable). In this case, more specific classifiers are generated by the particular specialization process. Thus, the oscillation of classifiers is at the heart of the model improvement process.

Specializing a classifier when it oscillates is a way to use the error of the prediction so as to improve the model, exactly as it is done in the context of ANN.

This way of learning is justified by the fact that both systems include a capacity of generalization in their models. Otherwise, it would be simpler just to include any new experience in the anticipatory model without having to encompass a prediction and correction process. The point is that the prediction can be general and the correction preserves this generality as much as it can. Interestingly, however, generalization is not exactly of the same nature in ANN and in ALCSs.

As a conclusion, both classes of systems exhibit a synergy between learning, prediction, and generalization, learning being used to improve general predictions, but also predictions being at the heart of learning general features of the environment.

6.3 Model Builders and non-Markov Problems

As explained in section 3.1, a non-Markov problem is a problem in which the current sensations of the animat are not always sufficient to choose the best action. In such problems, the animat must incorporate an internal state model representation M^S providing a further source of information for choosing the best action. The information in question generally comes from the more or less immediate past of the animat. An animat which does not incorporate such an internal state model is said to be "reactive". Reactive animats cannot behave optimally in non-Markov problems.

In order to prevent misinterpretations, we must warn the reader about the fact that an internal state model differs from an internal predictive model. In fact, an internal predictive model alone does not enable the animat to behave optimally in a non-Markov problem. Rather than information about the immediate past of the animat, predictive models only provide information about the "atemporal" structure of the problem (that is, information about the possible future). In particular, if the animat has no means to disambiguate aliased perceptions, it will build an aliased model. Thus an animat can be both reactive, that is, unable to behave optimally in non-Markov environments, and explicitly anticipatory, that is, able to build a predictive model of this environment and bias its action decisions on future predictions, without solving the non-Markov problem.

7 Conclusion

This overview of internal models and anticipatory behavior showed that a lot of future research is needed to understand exactly when which anticipations are useful or sometimes even mandatory in an environment to yield competent adaptive behavior. Although psychological research proves that anticipatory behavior takes place in at least higher animals, a clear understanding of the *how*, the *when*, and the *which* is not available. Thus, one essential direction of future research is to identify environmental characteristics in which distinct anticipatory mechanisms are helpful or necessary.

Several more concrete research directions can be suggested. (1) It seems important to quantify when anticipatory behavior can be adapted faster than stimulus-response behavior. For example, in a dynamic environment some predictive knowledge may be assumed to be stable so that behavior can be adapted by the means of this knowledge. (2) It appears interesting to investigate how to balance reactive and anticipatory mechanisms and how to allow a proper interaction. A proper architecture of motivations and emotions might play an important role in this respect. (3) Adaptive mechanisms that are initially anticipatory and then become short circuited reactive demand further research effort. For example, initial hard practice of playing an instrument becomes more and more automatic and is eventually only guided by a correct feeling of its functioning. Can we create a similar adaptive motor-control mechanism? (4) The functioning of attentional processes influenced by sensory anticipations needs to be investigated further. When are such attentional mechanisms beneficial, when does the drawback due to inattentional blindness effects overshadow the benefits? (5) The benefit of simulating intentions and behavior of other animats requires further research effort. Which processes are necessary to create beneficial social relationships? Which mechanisms can result in mutual benefit, which mechanisms can cause unilateral benefit?

This small but broad list shows that future work in anticipatory learning systems promises fruitful research projects and new exciting insights in the field of adaptive behavior. We hope that our overview of current insights in anticipatory mechanisms and the available systems provide a basis for future research efforts. Moreover, we want to encourage the development of the distinctions between anticipatory behavior mechanisms. While implicit and payoff anticipatory mechanisms appear to be rather clear cut, sensory and state anticipatory behavior comprise many different forms and mechanisms. Future research will show which characteristics should be used to distinguish the different mechanisms further.

Acknowledgments

The authors would like to thank Stewart Wilson, Joanna Bryson, and Mark Witkowski for useful comments on an earlier draft of this introduction.

This work was funded by the German Research Foundation (DFG) under grant HO1301/4-3.

References

1. Arbib, M.: The mirror system, imitation, and the evolution of language. In Dautenhahn, K., Nehaniv, C.L., eds.: Imitation in animals and artifacts. MIT Press, Cambridge, MA (2002)
2. Baluja, S., Pomerleau, D.A.: Using the representation in a neural network's hidden layer for task-specific focus on attention. Proceedings of the Fourteenth International Joint Conference on Artificial Intelligence (1995) 133–141
3. Baluja, S., Pomerleau, D.A.: Expectation-based selective attention for visual monitoring and control of a robot vehicle. Robotics and Autonomous Systems **22** (1997) 329–344
4. Barto, A.G., Mahadevan, S.: Recent advances in hierarchical reinforcement learning. Discrete event systems (2003, to appear)
5. Bellman, R.E.: Dynamic programming. Princeton University Press, Princeton, NJ (1957)
6. Booker, L., Goldberg, D.E., Holland, J.H.: Classifier systems and genetic algorithms. Artificial Intelligence **40** (1989) 235–282
7. Brooks, R.A.: Intelligence without reason. Proceedings of the 12th International Joint Conference on Artificial Intelligence (1991) 569–595
8. Butz, M.V.: Anticipatory learning classifier systems. Kluwer Academic Publishers, Boston, MA (2002)
9. Cassandra, A.R., Kaelbling, L.P., Littman, M.L.: Acting optimally in partially observable stochastic domains. Proceedings of the Twelfth National Conference on AI (1994) 1023–1028
10. Davidsson, P.: Learning by linear anticipation in multi-agent systems. In Weiss, G., ed.: Distributed artificial intelligence meets machine learning, Berlin Heidelberg, Springer-Verlag (1997) 62–72
11. Drescher, G.L.: Made-up minds, a constructivist approach to artificial intelligence. MIT Press, Cambridge, MA (1991)
12. Elman, J.L.: Finding structure in time. Cognitive Science **14** (1990) 179–211
13. Fikes, R.E., Nilsson, N.J.: STRIPS: A new approach to the application of theorem proving to problem solving. Artificial Intelligence **2** (1971) 189–208
14. Gallese, V.: The 'shared manifold' hypothesis: From mirror neurons to empathy. Journal of Consciousness Studies: Between Ourselves - Second-Person Issues in the Study of Consciousness **8** (2001) 33–50
15. Gallese, V., Goldman, A.: Mirror neurons and the simulation theory of mindreading. Trends in Cognitive Sciences **2** (1998) 493–501
16. Gérard, P., Meyer, J.A., Sigaud, O.: Combining latent learning and dynamic programming in MACS. European Journal of Operational Research (submitted, 2003)
17. Gérard, P., Stolzmann, W., Sigaud, O.: YACS: A new learning classifier system with anticipation. Soft Computing **6** (2002) 216–228
18. Gérard, P., Sigaud, O.: Adding a generalization mechanism to YACS. Proceedings of the Genetic and Evolutionary Computation Conference (GECCO-2001) (2001) 951–957
19. Gérard, P., Sigaud, O.: YACS: Combining dynamic programming with generalization in classifier systems. In Lanzi, P.L., Stolzmann, W., Wilson, S.W., eds.: Advances in learning classifier systems: Third international workshop, IWLCS 2000. Springer-Verlag, Berlin Heidelberg (2001) 52–69
20. Goldberg, D.E.: Genetic algorithms in search, optimization and machine learning. Addison-Wesley, Reading, MA (1989)

21. Herbart, J.: Psychologie als Wissenschaft neu gegründet auf Erfahrung, Metaphysik und Mathematik. Zweiter, analytischer Teil. August Wilhem Unzer, Königsberg, Germany (1825)
22. Hoffmann, J., Sebald, A., Stöcker, C.: Irrelevant response effects improve serial learning in serial reaction time tasks. Journal of Experimental Psychology: Learning, Memory, and Cognition **27** (2001) 470–482
23. Holland, J.H.: Adaptation in natural and artificial systems. The University of Michigan Press (1975)
24. Holland, J.H., Holyoak, K.J., Nisbett, R.E., Thagard, P.R.: Induction. MIT Press (1986)
25. Holland, J.H., Reitman, J.S.: Cognitive systems based on adaptive algorithms. Pattern Directed Inference Systems **7** (1978) 125–149
26. Holland, J.H.: Properties of the bucket brigade algorithm. Proceedings of an International Conference on Genetic Algorithms and their Applications (1985) 1–7
27. Kaelbing, L.P., Littman, M.L., Moore, A.W.: Reinforcement learning: A survey. Journal of Artificial Intelligence Research **4** (1996) 237–285
28. Koch, C., Ullmann, S.: Shifts in selective attention: Towards the underlying neural circuitry. Human Neurobiology **4** (1985) 219–227
29. Kunde, W.: Response-effect compatibility in manual choice reaction tasks. Journal of Experimental Psychology: Human Perception and Performance **27** (2001) 387–394
30. Kuvayev, L., Sutton, R.S.: Model-based reinforcement learning with an approximate, learned model. In: Proceedings of the ninth yale workshop on adaptive and learning systems, New Haven, CT (1996) 101–105
31. LaBerge, D.: Attentional processing, the brain's art of mindfulness. Harvard University Press, Cambridge, MA (1995)
32. Lanzi, P.L.: An analysis of generalization in the XCS classifier system. Evolutionary Computation **7** (1999) 125–149
33. Lanzi, P.L.: Learning classifier systems from a reinforcement learning perspective. Soft Computing **6** (2002) 162–170
34. Mack, A., Rock, I.: Inattentinal blindness. MIT Press (Cambridge, MA)
35. Moore, A.W., Atkeson, C.: Prioritized sweeping: Reinforcement learning with less data and less real time. Machine Learning **13** (1993) 103–130
36. Newell, A., Simon, H.A., Shaw, J.C.: Elements of a theory of human problem solving. Psychological Review **65** (1958) 151–166
37. Pashler, H., Johnston, J.C., Ruthruff, E.: Attention and performance. Annual Review of Psychology **52** (2001) 629–651
38. Pashler, H.E.: The psychology of attention. MIT Press, Cambridge, MA (1998)
39. Pavlov, I.P.: Conditioned reflexes. London: Oxford (1927)
40. Peng, J., Williams, R.J.: Efficient learning and planning within the dyna framework. Adaptive Behavior **1** (1993) 437–454
41. Rizzolatti, G., Fadiga, L., Gallese, V., Fogassi, L.: Premotor cortex and the recognition of motor actions. Cognitive Brain Research **3** (1996) 131–141
42. Rummery, G.A., Niranjan, M.: On-line Q-learning using connectionist systems. Technical Report CUED/F-INFENG/TR 166, Engineering Department, Cambridge University (1994)
43. Schubotz, R.I., von Cramon, D.Y.: Functional organization of the lateral premotor cortex. fMRI reveals different regions activated by anticipation of object properties, location and speed. Cognitive Brain Research **11** (2001) 97–112
44. Simons, D.J., Chabris, C.F.: Gorillas in our midst: Sustained inattentional blindness for dynamic events. Perception **28** (1999) 1059–1074

45. Sjölander, S.: Some cognitive break-throughs in the evolution of cognition and consciousness, and their impact on the biology language. Evolution and Cognition **1** (1995) 3–11
46. Skinner, B.F.: The behavior of organisms. Appleton-Century Crofts, Inc., New-York (1938)
47. Skinner, B.F.: Beyond freedom and dignity. Bantam/Vintage, New York (1971)
48. Stock, A., Hoffmann, J.: Intentional fixation of behavioral learning or how R-E learning blocks S-R learning. European Journal of Cognitive Psychology (2002) in press.
49. Stolzmann, W.: Antizipative Classifier Systems [Anticipatory classifier systems]. Shaker Verlag, Aachen, Germany (1997)
50. Stolzmann, W.: Anticipatory classifier systems. Genetic Programming 1998: Proceedings of the Third Annual Conference (1998) 658–664
51. Stolzmann, W., Butz, M.V., Hoffmann, J., Goldberg, D.E.: First cognitive capabilities in the anticipatory classifier system. From Animals to Animats 6: Proceedings of the Sixth International Conference on Simulation of Adaptive Behavior (2000) 287–296
52. Sutton, R.S., Barto, A.G.: Reinforcement learning: An introduction. MIT Press (1998)
53. Sutton, R.: Reinforcement learning architectures for animats. From animals to animats: Proceedings of the First International Conference on Simulation of Adaptative Behavior (1991)
54. Sutton, R., Precup, D., Singh, S.: Between MDPs and semi-MDPs: A framework for temporal abstraction in reinforcement learning. Artificial Intelligence **112** (1999) 181–211
55. Tani, J.: Model-based learning for mobile robot navigation from the dynamical system perspective. IEEE Transactions on System, Man and Cybernetics **26** (1996) 421–436
56. Tani, J.: An interpretation of the "self" from the dynamical systems perspective: A constructivist approach. Journal of Consciousness Studies **5** (1998) 516–542
57. Tani, J.: Learning to perceive the world as articulated: An approach for hierarchical learning in sensory-motor systems. Neural Networks **12** (1999) 1131–1141
58. Thistlethwaite, D.: A critical review of latent learning and related experiments. Psychological Bulletin **48** (1951) 97–129
59. Thompson, E.: Empathy and consciousness. Journal of Consciousness Studies: Between Ourselves - Second-Person Issues in the Study of Consciousness **8** (2001) 1–32
60. Thorndike, E.L.: Animal intelligence: Experimental studies. Macmillan, New York (1911)
61. Tolman, E.C.: Purposive behavior in animals and men. Appletown, New York (1932)
62. Tolman, E.C.: The determiners of behavior at a choice point. Psychological Review **45** (1938) 1–41
63. Tolman, E.C.: Cognitive maps in rats and men. Psychological Review **55** (1948) 189–208
64. Tolman, E.C.: Principles of purposive behavior. In Koch, S., ed.: Psychology: A study of science, New York, McGraw-Hill (1959) 92–157
65. Watkins, C.J.: Learning with delayed rewards. PhD thesis, Psychology Department, University of Cambridge, England (1989)
66. Wilson, S.W.: ZCS, a zeroth level classifier system. Evolutionary Computation **2** (1994) 1–18

67. Wilson, S.W.: Classifier fitness based on accuracy. Evolutionary Computation **3** (1995) 149–175
68. Wilson, S.W.: Mining oblique data with XCS. In Lanzi, P.L., Stolzmann, W., Wilson, S.W., eds.: Advances in learning classifier systems: Third international workshop, IWLCS 2000, Berlin Heidelberg, Springer-Verlag (2001)
69. Wilson, S.W.: Knowledge growth in an artificial animal. In Grefenstette, J.J., ed.: Proceedings of an international conference on genetic algorithms and their applications, Carnegie-Mellon University, Pittsburgh, PA (1985) 16–23
70. Witkowski, C.M.: Schemes for learning and behaviour: A new expectancy model. PhD thesis, Department of Computer Science, University of London, England (1997)
71. Wolpert, D.H.: The lack of a priori distinctions between learning algorithms. Neural Computation **8** (1995) 1341–1390