Internal Models and Anticipations in Adaptive Learning Systems

Martin V. Butz[2,3], Olivier Sigaud[1], and Pierre Gerard[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 workshop Adaptive Behavior in Anticipatory Learning Systems 2002 (ABiALS 2002) held in association with the seventh international conference on the Simulation of Adaptive Behavior (SAB 2002) in Edinburgh, Scotland, is the first of its kind. The explicit investigation of anticipations in relation to adaptive behavior is a recent approach. In this introduction to the workshop, we first provide psychological background that motivates and inspires the study of anticipations in the adaptive behavior field. Next, we endow the workshop with a basic framework for the study of anticipations in adaptive behavior. Different anticipatory mechanisms are identified and categorized. First fundamental distinctions are drawn between implicitly anticipatory mechanisms, payoff anticipations, sensorial anticipations, and state anticipations. A case study allows further insights into the drawn distinctions. Moreover, an overview of all accepted workshop contributions is provided categorizing the contributions in the light of the outlined distinctions and highlighting the relations to each other. Many future research direction are suggested.

1 Introduction

The idea that *anticipations*, referring to the influence of predictions on actual behavior, 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.

In cognitive psychology anticipations have been experimentally shown to influence behavior ranging from simple reaction time tasks to elaborate reasoning tasks (Kunde, 2001; Stock & Hoffmann, 2002). It becomes more and more obvious that anticipations influence actual behavior as well as memory mechanisms and attention (Pashler, Johnston, & Ruthruff, 2001). 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 (Schubotz & von Cramon, 2001).

In animal behavior studies the observation of *latent learning* was the first indicator for anticipatory behavior. 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. Tolman, 1932; Thistlethwaite, 1951). Similarly, in more recent *outcome devaluation* experiments (e.g. Adams & Dickinson, 1981; Colwill & Rescorla, 1985) animals show that they have learned not only the quality of an action in a particular situation but also the actual effect of that action.

Although it might be true that over all constructible learning problems any learning mechanisms can generalize its knowledge as good, or as bad, as any other one (Wolpert, 1995), these 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* (Wilson, 1985), the incorporation of anticipatory mechanisms seems mandatory.

This workshop is dedicated to the study of possible anticipatory mechanisms. On the one hand, we are interested in *how* anticipatory mechanisms can be incorporated in animats, that is, which structures are necessary for anticipatory behavior systems. On the other hand, we are interested in *when* anticipatory mechanisms are actually helpful in animats, that is, which environmental preconditions favor anticipatory behavior. This first workshop on anticipatory adaptive behavior does not provide exact answers to any of those questions, but it certainly provides many ideas and insights that promise to eventually lead to concrete answers.

To approach the *how* and *when*, it is necessary to distinguish 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 paper, we draw a first distinction between (1) *implicitly 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) *sensorial anticipatory* mechanisms in which future predictions influence sensorial (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 paper is structured as follows. First, psychology's knowledge about anticipations is sketched out. Moreover, the general framework of anticipatory behavioral control is introduced. 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 future research efforts. Section 7 briefly introduces all ABiALS 2002 contributions and discusses their relations to each other as well as to the suggested distinct anticipatory mechanisms. The conclusions outline many diverse future research directions tied to the study of anticipations in adaptive behavior.

2 The background from psychological research

In order to motivate the usage of anticipations in adaptive behavior research, this section provides background from cognitive psychology. Moreover, Hoffmann's learning theory of anticipatory behavioral control is introduced as a first anticipatory learning and behavioral framework. In Witkowski (2002), further relevant psychological background can be found.

2.1 Behaviorism versus Tolman's Expectancy Model

Early suggestions of anticipations in behavior date back to Herbart (1825). 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 denying any sort of mental state representation. Two of the predominant principles in the behaviorist world were *classical conditioning* and *operant conditioning*.

Pavlov first introduced classical conditioning (Pavlov, 1927). 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 (1911) 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".

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.

In sharp contrast with these behaviorist theories, Tolman (Tolman, 1932; Tolman, 1938; Tolman, 1948) proposed that, additionally to conditioned learning, *latent learning* takes place in animals. He showed that animals learn a sort of internal representation of the world independent of any reinforcement. 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). Seward (1949) provides one of the soundest experiments in this fashion. The observation of latent learning led Tolman to propose that animals do 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.(Tolman, 1959, 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 used for anticipatory behavior.

2.2 Hoffmann's Anticipatory Behavioral Control

Hoffmann's framework of *anticipatory behavioral control* supposes how such expectancies might be learned and how they can influence actual behavior. The key insight of Hoffmann's position is expressed in the following quotation:

> [...] that first, purposive behavior (R) is always accompanied by anticipations of the effects (E_{ant}) expected according to the previous experience in the situation (S). Secondly, it is assumed that the anticipation of (E_{ant}) is continually compared to the real effect (E_{real}). Correct anticipations should increase the behavior-related bond between the confirmed anticipations and the stimuli of the situation from which the behavior emerged. [...] Incorrect anticipations should cause a differentiation of the situation-related conditions with respect to the associated behavioral consequences.(Hoffmann, 1993, p.45, own translation)

In the context of animats, the proposition can be defined as follows. An anticipatory animat always forms predictions of expected, possibly action dependent, outcomes according to its predictive model before action execution. The model is modified by comparing the predictions with the real results focusing primarily on action-effect contingencies. Secondarily, conditional dependencies are taken into account.

This line of psychological research had an important influence in computer modeling since both Stolzmann and Butz (Stolzmann, 1997; Stolzmann, 1998; Butz, 2002) are coming from that school and propose Learning Classifier System (LCS) models directly inspired from this view. In the same line of research, albeit with a less direct influence, Witkowski (1997) and Gérard and Sigaud (2001b) have designed algorithms relying on very similar ideas.

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 (Wilson, 1985). Research in artificial intelligence moved away from the traditional predicate logic and planning approaches to intelligence without representation (Brooks, 1991). The main idea is that intelligent behavior can arise without any high-level cognition. Smart connections from sensors to actions 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. Researchers in adaptive behavior and artificial life increasingly develop hybrid approaches. The embodied intelligent agents are 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 or to handle unforeseen situations. Essentially, the agent is able to learn and draw inferences by the means of internal representations and mechanisms.

The cognitive mechanisms that are 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 In this workshop, we are mainly 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;

- $O : X \to \Pi(Y)$ the observation function, where $\Pi(Y)$ is the set of all probability distributions over Y;
- $R : X \times U \times X \to \Re^r$ 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 so called "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 $P(t) \in \Re$. 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 partly 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 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, sensorial anticipations, and state anticipations. All four types of anticipations 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. Sensorial 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.

In nature, even if a life-form behaves purely reactive, 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 implicitly anticipatory 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 implicitly 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. This workshop certainly does not address all algorithms imaginable in this general framework (basically at least the whole artificial intelligence field) but focuses on the more explicit types of anticipations outlined below. However, it is necessary to understand the difference between such implicitly anticipatory animats and animats that form some kind of explicit prediction that influences 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 animat. 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 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 (Rummery & Niranjan, 1994; Sutton & Barto, 1998), and off-policy RL algorithms, such as Q-learning (Watkins, 1989; Sutton & Barto, 1998).

Sensorial Anticipations While in payoff anticipations predictions are restricted to payoff, in sensorial anticipations predictions are unrestricted. However, sensorial 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 such predictions, the animat must use a (not necessarily complete) predictive model M^P of its environment. 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.

Sensory anticipations strongly relate to preparatory attention in psychology (LaBerge, 1995; Pashler, 1998) in which top-down processes such as task-related expectations influence sensory processing and learning. Thus, behavior is not directly influenced by future predictions, but the predictions influence sensorial (pre-)processing. The (pre-)processing then possibly influences behavior. Also learning might be influenced by such sensorial anticipations as suggested in psychological studies on learning (Hoffmann, Sebald, & Stöcker, 2001; Stock & Hoffmann, 2002).

State Anticipations Maybe the most interesting group of anticipations is the one in which the animat forms explicit predictions about future states that influence current decision making. As in sensorial anticipations, a predictive model must be available to the animat or it must be learned by the animat. In difference to sensorial anticipations, however, state anticipations directly influence current behavioral decision making. An explicit anticipatory animat is visualized in figure 1. The essential property is that prediction(s) about future state(s) influence the actual action decision.

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* (Bellman, 1957). The most extreme cases of such high-level planning approaches can be found in early artificial intelligence work such as the general problem solver (Newell, Simon, & Shaw, 1958) or the STRIPS language (Fikes & Nilsson, 1971). Nowadays, somewhat related approaches try to focus on *local mechanisms* that extract only *relevant environmental information.*

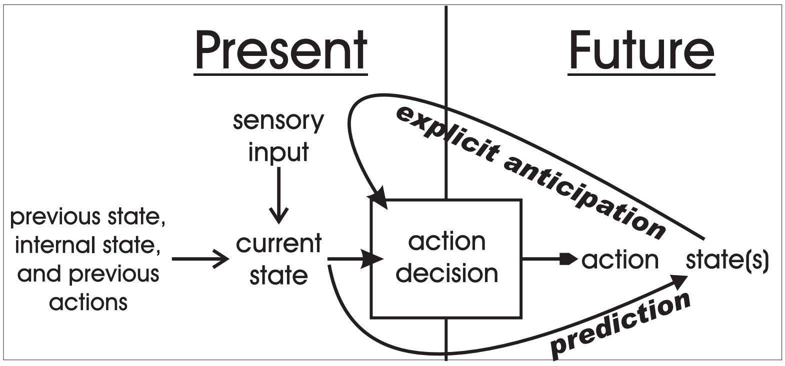

Fig. 1. Explicit anticipations influence the actual action decision by predictions about the future.

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 predictions determine behavior.

Further distinctions in anticipatory animats that use a predictive model for decision making are evident in the structure of the model representation, the completeness of the model representation, the information the model representation is based on, and the learning and generalization mechanisms that may change the model over time. 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. Finally, learning and generalization mechanisms give rise to further crucial differences in the availability, the efficiency, and the utility of the predictive model.

With a proper definition of animats and four fundamental classes of anticipations 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 (Kaelbing, Littman, & Moore, 1996; Sutton & Barto, 1998) 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 Cassandra, Kaelbling, & Littman, 1994). 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 Watkins (1989). 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 assures the generation of 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. Approaches that generalize 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 Holland and Reitman (1978). 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 (Holland, 1975; Goldberg, 1989). 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 (Lanzi, 2002). 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. However, the generalization over the perceptions promises the faster adaptation in dynamic environments as well as the more compact policy representation in an environment 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 (Holland, 1985) to resemble the Q-learning mechanism propagating Q-values over the population of classifiers (Wilson, 1994; Wilson, 1995). Moreover, Wilson drastically simplified the LCS model (Wilson, 1994). Then, he modified Holland's original strength-based criterion for learning — the more a rule receives reward (on average), the more fit it is (Holland, 1975; Holland, Holyoak, Nisbett, & Thagard, 1986; Booker, Goldberg, & Holland, 1989) — by a new criterion relying on the accuracy of the reward prediction of each rule (Wilson, 1995). 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 sensorial anticipatory or state anticipatory. Important differences of the systems are outlined below.

5.1 Model-based Reinforcement Learning

The dynamical architecture *Dyna* (Sutton, 1991) 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 (Moore & Atkeson, 1993; Peng & Williams, 1993) or adopting the mechanism to a tile coding approach (Kuvayev & Sutton, 1996). 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 Drescher (1991). 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 (1997) 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) (Stolzmann, 1998; Butz, 2002; Gérard & Sigaud, 2001b; Gérard, Stolzmann, & Sigaud, 2002) 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 (Stolzmann, 1997; Stolzmann, 1998) and was later extended to its current state of the art, ACS2 (Butz, 2002). 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 (Butz, 2002). 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 (Stolzmann, 1998; Stolzmann, Butz, Hoffmann, & Goldberg, 2000; Butz, 2002).

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 (Gérard, Stolzmann, & Sigaud, 2002; Gérard & Sigaud, 2001a). 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 (Gérard, Meyer, & Sigaud, 2002) 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 thus 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 (Tani, 1996; Tani, 1999). 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 (Baluja & Pomerleau, 1995; Baluja & Pomerleau, 1997). The mechanism is based on the ideas of visual attention modeling in Koch and Ullmann (1985). 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. Sensorial 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 (Tani, 1996). 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 (Elman, 1990). 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 and thus 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 correctly the next inputs helps stabilizing the behavior of its agents and, more generally, that using anticipations results in a bipolarization 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 (Tani, 1998), 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 Koch and Ullmann (1985), 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 sensorial anticipatory mechanisms that influence attentional mechanisms similar to Baluja's visual attention mechanism but embedded in a bigger modular structure.

In Tani (1999), 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 with 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 Davidsson (1997). 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 (Bellman, 1957) and it is adapted to the RL framework in the Dyna architecture (Sutton, 1991; Sutton & Barto, 1998). 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 sensori-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 (Wilson, 2001) 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 (Lanzi, 1999). 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 Overview of ABiALS 2002 Contributions

As can be inferred from the study above, anticipations can be understood in many different forms and, in the broadest sense, anticipations can be found everywhere. The major aim of ABiALS 2002 was to study anticipations by the means of a predictive model but the discussion of anticipatory mechanisms in the broader sense was encouraged, as well. We hope that the available contributions strengthen the understanding of the drawn distinction between implicitly-anticipatory, payoff-anticipatory, sensorial-anticipatory, and state-anticipatory animats. Moreover, we hope that the contributions provide interesting points of departure for future research on anticipations in adaptive behavior.

A basic framework for an incorporation of model-based anticipatory processes can be found in the Behavior-Oriented Design (BOD) approach presented by Bryson (Bryson, 2002). In the BOD framework, the most useful achievements of Behavior-Based Artificial Intelligence (BBAI) and Multi-Agent Systems (MAS) are combined. While both approaches study intelligence arising from simple somewhat interacting modules (or agents), the two fields have somewhat ignored each other's research. BOD is a design proposition that combines the two directions. It enables the designer of an intelligent animat to provide as much expert knowledge as available stressing high modularity. The framework promises to facilitate the incorporation of model-based anticipations in the form of modular expectations and predictions that influence the action selection process.

Three of the remaining six contributions deal with state-anticipatory animats in which the predictive model is represented by a set of predictive rules (Baldassarre, 2002; Witkowski, 2002; Butz & Goldberg, 2002). Baldassarre extends Sutton's Dyna-PI model (Sutton, 1991) with additional goal representations (the "matcher") and goal dependent planning algorithms. Furthermore, internal planning steps are not executed in fixed intervals, but depend on the confidence of the animat of reaching a given goal from the current state. This confidence measure reflects the animat's belief in its own predictions and results in a controlled "thinking before acting". Baldassarre also provides many interesting parallels to the current knowledge of the neural structure and mechanisms underlying human planning.

Witkowski provides further cognitive psychology background and approaches explicit anticipations in a more general framework. He distinguishes four essential capabilities for anticipatory animats: (1) action independent future predictions; (2) action dependent future predictions; (3) reinforcement independent action ranking; and (4) guided structural learning by detecting unpredicted events (that is, biased learning of a predictive model). Next, he develops a general anticipatory framework that comprises reactive behavior and learning, conditioning, goal propagation, and goal oriented behavior. Finally, Witkowski explains his anticipatory animat program SRS/E (see also section 5.3) in his Dynamic Expectancy Model based on the developed anticipatory framework.

Butz and Goldberg enhance the anticipatory classifier system ACS2 with further state-anticipatory mechanisms. The paper addresses the online generalization of state values while learning a predictive model. State values reflect the utility of reaching a state given a current problem (in the form of a POMDP). For ungeneralized states, the values are identical to values that can be determined by the dynamic programming algorithm approximating the Bellman equation (Bellman, 1957). The resulting system, XACS, implements a predictive model learning module and a separate reinforcement learning module generalizing the representations of both modules online. Behavior is state-anticipatory in that future predictions and the values of those predicted states determine actual behavior. The interaction of multiple reinforcement modules is suggested.

Although starting from the consequence driven systems theory, Bozinovski's contribution (Bozinovski, 2002) provides a possibility of how such multiple reinforcement modules could be combined. The question is addressed, what motivation and what emotion are in an anticipatory system. Motivations for anticipatory behavior are characterized by the anticipation of future emotional consequences. Motivations are represented in motivational graphs. Emotions coordinate action tendencies in current states as well as influence motivations due to the anticipated emotional values of future states. Emotions are represented in emotional polynoms. The paper is a great contribution to the understanding of importance and influence of anticipations in an emotional adaptive system. In the light of our distinction of anticipatory animats, Bozinovski suggests a state-anticipatory animat in which predictions do not directly influence action decisions but behavior is influenced indirectly by the means of the emotional polynoms that assign weights to predictions and by the means of motivational graphs that prioritize the predicted emotions.

Laaksolahti and Boman (2002) address the usage of predictions in the guidance of the plot of a story. A system is suggested that uses future predictions to prevent undesired states and to promote the sought plot. It is suggested to implement a God-like mechanism that forms predictions that possibly cause the alteration of the current states of the story. For example, it is suggested to prevent undesired story outcomes by altering the emotional states of persons in the story by such mechanisms. Similar to Bozinovski's emotional framework, the guidance of plot can be classified as a state anticipatory system that does not directly influence behavior but it influences behavior indirectly by altering behavior-relevant parameters, such as emotional states.

The final contribution addresses the value of predictions in artificial stock markets (Edmonds, 2002). Edmonds uses two genetic programming modules, one used to learn to trade, the other one trying to predict the future behavior of the market.

Interestingly, the predictions are considered as extra inputs that may or may not be used by the module learning to trade. Somewhat counter-intuitively, the presented results do not indicate that the additional information always improves performance. Two characteristics of an environment are mentioned that seem to be essential for a benefit from predictive model information: (1) the prediction must be feasible; (2) some advantage can be drawn from this predictive information. It is further speculated that prediction might be useful in the short run. In the long run, though, the explicit anticipatory behavior becomes directly built into the reactions of the animat.

8 Conclusion

This introduction to the first workshop on adaptive behavior in anticipatory learning systems (ABiALS 2002) shows that a lot of future research is needed to understand exactly when and which anticipations are useful or sometimes even mandatory in an environment to yield competent adaptive behavior. Although psychological research proves that anticipations take 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.

On the learning side, it seems to be important to identify when anticipatory learning is actually faster than stimulus-response learning. Moreover, it appears interesting to investigate how to balance the two learning methods and how to allow a proper interaction. Finally, the apparent step by step adaptation from anticipatory behavior to short circuited reactive, or at least non-anticipatory, behavior requires further research effort.

The provided first step to the relation of motivations, emotions, and anticipations in ABiALS 2002 also promises fruitful future research. The concepts also appear to relate to the behavioral shift from short-term beneficial model-based anticipatory behavior to long-term beneficial reactive-based behavior. Initial hard practice of any task becomes more and more automatic and is eventually only guided by a correct feeling of its functioning.

Further issues need to be addressed in the anticipatory framework such as the general functioning of attentional processes with sensorial anticipations, the simulation of intentions and behavior of other animats, or the interaction of reactive and anticipatory behavior. This small but broad list shows that future work in anticipatory learning systems promises fruitful research and new exciting insights in the adaptive behavior field and in the general quest for designing competent and autonomous adaptive animats.

Acknowledgments

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

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

References

Adams, C., & Dickinson, A. (1981). Instrumental responding following reinforcer devaluation. *Quarterly Journal of Experimental Psychology*, *33*(B), 109–121.

Baldassarre, G. (2002). A biologically plausible model of human planning based on neural networks and Dyna-PI models. In Butz, M. V., Gérard, P., & Sigaud, O. (Eds.), *Adaptive Behavior in Anticipatory Learning Systems (ABIALS'02)* Edinburgh, Scotland.

Baluja, S., & Pomerleau, D. A. (1995). Using the representation in a neural network's hidden layer for task-specific focus on attention. In *Proceedings of the Fourteenth International Joint Conference on Artificial Intelligence* pp. 133–141. San Francisco, CA: Morgan Kaufmann.

Baluja, S., & Pomerleau, D. A. (1997). Expectation-based selective attention for visual monitoring and control of a robot vehicle. *Robotics and Autonomous Systems*, *22*, 329–344.

Bellman, R. E. (1957). *Dynamic Programming.* Princeton, NJ: Princeton University Press.

Booker, L., Goldberg, D. E., & Holland, J. H. (1989). Classifier systems and genetic algorithms. *Artificial Intelligence*, *40*(1-3), 235–282.

Bozinovski, S. (2002). Motivation and emotion in anticipatory behavior of consequence driven systems. In Butz, M. V., Gérard, P., & Sigaud, O. (Eds.), *Adaptive Behavior in Anticipatory Learning Systems (ABIALS'02)* Edinburgh, Scotland.

Brooks, R. A. (1991). Intelligence without reason. In Myopoulos, John; Reiter, R. (Ed.), *Proceedings of the 12th International Joint Conference on Artificial Intelligence* pp. 569–595. Sydney, Australia: Morgan Kaufmann.

Bryson, J. J. (2002). Reexamining behavior-based artificial intelligence. In Butz, M. V., Gérard, P., & Sigaud, O. (Eds.), *Adaptive Behavior in Anticipatory Learning Systems (ABIALS'02)* Edinburgh, Scotland.

Butz, M. V. (2002). *Anticipatory learning classifier systems.* Genetic Algorithms and Evolutionary Computation. Boston, MA: Kluwer Academic Publishers.

Butz, M. V., & Goldberg, D. E. (2002). Generalized state values in an anticipatory learning classifier system. In Butz, M. V., Gérard, P., & Sigaud, O. (Eds.), *Adaptive Behavior in Anticipatory Learning Systems (ABiALS'02)* Edinburgh, Scotland.

Cassandra, A. R., Kaelbling, L. P., & Littman, M. L. (1994). Acting optimally in partially observable stochastic domains. *Proceedings of the Twelfth National Conference on AI*, 1023–1028.

Colwill, R. M., & Rescorla, R. A. (1985). Postconditioning devaluation of a reinforcer affects instrumental learning. *Journal of Experimental Psychology: Animal Behavior Processes*, *11*(1), 120–132.

Davidsson, P. (1997). Learning by linear anticipation in multi-agent systems. In Weiss, G. (Ed.), *Distributed Artificial Intelligence Meets Machine Learning* pp. 62–72. Berlin Heidelberg: Springer-Verlag.

Drescher, G. L. (1991). *Made-Up Minds, a constructivist approach to artificial intelligence.* Cambridge, MA: MIT Press.

Edmonds, B. (2002). Exploring the value of prediction in an artificial stock market. In Butz, M. V., Gérard, P., & Sigaud, O. (Eds.), *Adaptive Behavior in Anticipatory Learning Systems (ABiALS'02)* Edinburgh, Scotland.

Elman, J. L. (1990). Finding structure in time. *Cognitive Science*, *14*, 179–211.

Fikes, R. E., & Nilsson, N. J. (1971). Strips: A new approach to the application of theorem proving to problem solving. *Artificial Intelligence*, *2*, 189–208.

Gérard, P., Meyer, J.-A., & Sigaud, O. (2002). Combining latent learning and dynamic programming in MACS. *European Journal of Operational Research*. submitted.

Gérard, P., & Sigaud, O. (2001a). Adding a generalization mechanism to YACS. In *Proceedings of the Genetic and Evolutionary Computation Conference (GECCO-2001)* pp. 951–957. San Francisco, CA: Morgan Kaufmann.

Gérard, P., & Sigaud, O. (2001b). YACS: Combining dynamic programming with generalization in classifier systems. In Lanzi, P. L., Stolzmann, W., & Wilson, S. W. (Eds.), *Advances in Learning Classifier Systems, LNAI 1996* pp. 52–69. Berlin Heidelberg: Springer-Verlag.

Gérard, P., Stolzmann, W., & Sigaud, O. (2002). YACS: a new Learning Classifier System with Anticipation. *Soft Computing*, *6*(3-4), 216–228.

Goldberg, D. E. (1989). *Genetic algorithms in search, optimization and machine learning.* Reading, MA: Addison-Wesley.

Herbart, J. (1825). *Psychologie als Wissenschaft neu gegründet auf Erfahrung, Metaphysik und Mathematik. Zweiter, analytischer Teil.* Koenigsberg, Germany: August Wilhem Unzer.

Hoffmann, J. (1993). *Vorhersage und erkenntnis [anticipation and cognition].* Hogrefe.

Hoffmann, J., Sebald, A., & Stöcker, C. (2001). Irrelevant response effects improve serial learning in serial reaction time tasks. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *27*, 470–482.

Holland, J. H. (1975). *Adaptation in Natural and Artificial Systems.* The University of Michigan Press.

Holland, J. H. (1985, july). Properties of the bucket brigade algorithm. In Grefenstette, J. J. (Ed.), *Proceedings of the 1st international Conference on Genetic Algorithms and their applications (ICGA85)* pp. 1–7. L.E. Associates.

Holland, J. H., Holyoak, K. J., Nisbett, R. E., & Thagard, P. R. (1986). *Induction.* MIT Press.

Holland, J. H., & Reitman, J. S. (1978). Cognitive Systems based on adaptive algorithms. *Pattern Directed Inference Systems*, *7*(2), 125–149.

Kaelbing, L. P., Littman, M. L., & Moore, A. W. (1996). Reinforcement Learning: A survey. *Journal of Artificial Intelligence Research*, *4*, 237–285.

Koch, C., & Ullmann, S. (1985). Shifts in selective attention: Towards the underlying neural circuitry. *Human Neurobiology*, *4*, 219–227.

Kunde, W. (2001). Response-effect compatibility in manual choice reaction tasks. *Journal of Experimental Psychology: Human Perception and Performance*, *27*, 387–394.

Kuvayev, L., & Sutton, R. S. (1996). Model-based reinforcement learning with an approximate, learned model. In *Proceedings of the Ninth Yale Workshop on Adaptive and Learning Systems* pp. 101–105. New Haven, CT.

Laaksolahti, J., & Boman, M. (2002). Anticipatory guidance of plot. In Butz, M. V., Gérard, P., & Sigaud, O. (Eds.), *Adaptive Behavior in Anticipatory Learning Systems (ABIALS'02)* Edinburgh, Scotland.

LaBerge, D. (1995). *Attentional processing, the brain's art of mindfulness.* Cambridge, MA: Harvard University Press.

Lanzi, P. L. (1999). An analysis of generalization in the XCS classifier system. *Evolutionary Computation*, *7*(2), 125–149.

Lanzi, P. L. (2002). Learning classifier systems from a reinforcement learning perspective. *Soft Computing*, *6*(3-4), 162–170.

Moore, A. W., & Atkeson, C. (1993). Prioritizes sweeping: Reinforcement learning with less data and less real time. *Machine Learning*, *13*, 103–130.

Newell, A., Simon, H. A., & Shaw, J. C. (1958). Elements of a theory of human problem solving. *Psychological Review*, *65*, 151–166.

Pashler, H., Johnston, J. C., & Ruthruff, E. (2001). Attention and performance. *Annual Review of Psychology*, *52*, 629–651.

Pashler, H. E. (1998). *The psychology of attention.* Cambridge, MA: MIT Press.

Pavlov, I. P. (1927). *Conditioned reflexes.* London: Oxford.

Peng, J., & Williams, R. J. (1993). Efficient learning and planning within the dyna framework. *Adaptive Behavior*, *1*(4), 437–454.

Rummery, G. A., & Niranjan, M. (1994). *On-line Q-learning using connectionist systems* (Technical Report CUED/F-INFENG/TR 166). Engineering Department, Cambridge University.

Schubotz, R. I., & von Cramon, D. Y. (2001). Functional organization of the lateral premotor cortex. fMRI reveals different regions activated by anticipation of object properties, location and speed. *Cognitive Brain Research*, *11*, 97–112.

Seward, J. P. (1949). An experimental analysis of latent learning. *Journal of Experimental Psychology*, *39*, 177–186.

Stock, A., & Hoffmann, J. (2002). Intentional fixation of behavioral learning or how R-E learning blocks S-R learning. *European Journal of Cognitive Psychology*. in press.

Stolzmann, W. (1997). *Antizipative Classifier Systems [Anticipatory classifier systems].* Aachen, Germany: Shaker Verlag.

Stolzmann, W. (1998). Anticipatory classifier systems. In Koza, J. R., Banzhaf, W., Chellapilla, K., Deb, K., Dorigo, M., Fogel, D., Grazon, M., Goldberg, D., Iba, H., & Riolo, R. (Eds.), *Genetic Programming 1998: Proceedings of the Third Annual Conference* (pp. 658–664). San Francisco, CA: Morgan Kaufmann.

Stolzmann, W., Butz, M. V., Hoffmann, J., & Goldberg, D. E. (2000). First cognitive capabilities in the anticipatory classifier system. In Meyer, J.-A., Berthoz, A., Floreano, D., Roitblat, H., & Wilson, S. W. (Eds.), *From Animals to Animats 6: Proceedings of the Sixth International Conference on Simulation of Adaptive Behavior* pp. 287–296. Cambridge, MA: MIT Press.

Sutton, R. (1991). Reinforcement learning architectures for animats. In Meyer, J. A., & Wilson, S. W. (Eds.), *From animals to animats: Proceedings of the First*

International Conference on Simulation of Adaptative Behavior Cambridge, MA: MIT Press.

Sutton, R. S., & Barto, A. G. (1998). *Reinforcement learning: An introduction.* MIT Press.

Tani, J. (1996). Model-based learning for mobile robot navigation from the dynamical system perspective. *IEEE Transactions on System, Man and Cybernetics*, *26*(3), 421–436.

Tani, J. (1998). An interpretation of the "self" from the dynamical systems perspective: A constructivist approach. *Journal of Consciousness Studies*, *5*(5-6), 516–542.

Tani, J. (1999). Learning to perceive the world as articulated: An approach for hierarchical learning in sensory-motor systems. *Neural Networks*, *12*, 1131–1141.

Thistlethwaite, D. (1951). A critical review of latent learning and related experiments. *Psychological Bulletin*, *48*(2), 97–129.

Thorndike, E. L. (1911). *Animal intelligence: Experimental studies.* New York: Macmillan.

Tolman, E. C. (1932). *Purposive behavior in animals and men.* New York: Appletown.

Tolman, E. C. (1938). The determiners of behavior at a choice point. *Psychological Review*, *45*(1), 1–41.

Tolman, E. C. (1948). Cognitive maps in rats and men. *Psychological Review*, *55*, 189–208.

Tolman, E. C. (1959). Principles of purposive behavior. In Koch, S. (Ed.), *Psychology: A study of science* pp. 92–157. New York: Mc<Graw-Hill.

Watkins, C. J. (1989). *Learning with delayed rewards.* Doctoral dissertation, Psychology Department, University of Cambridge, England.

Wilson, S. W. (1985). Knowledge growth in an artificial animal. In *Proceedings of an international conference on genetic algorithms and their applications* pp. 16–23. Carnegie-Mellon University, Pittsburgh, PA: John J. Grefenstette.

Wilson, S. W. (1994). ZCS, a Zeroth level Classifier System. *Evolutionary Computation*, *2*(1), 1–18.

Wilson, S. W. (1995). Classifier Fitness Based on Accuracy. *Evolutionary Computation*, *3*(2), 149–175.

Wilson, S. W. (2001). Mining oblique data with XCS. In Lanzi, P. L., Stolzmann, W., & Wilson, S. W. (Eds.), *Advances in Learning Classifier Systems: Proceedings of the Third International Workshop, LNAI 1996* Berlin Heidelberg: Springer-Verlag.

Witkowski, C. M. (1997). *Schemes for learning and behaviour: A new expectancy model.* Doctoral dissertation, Department of Computer Science, University of London, England.

Witkowski, C. M. (2002). Anticipatory learning: The animat as discovery engine. In Butz, M. V., Gérard, P., & Sigaud, O. (Eds.), *Adaptive Behavior in Anticipatory Learning Systems (ABiALS'02)* Edinburgh, Scotland.

Wolpert, D. H. (1995). The lack of a priori distinctions between learning algorithms. *Neural Computation*, *8*(7), 1341–1390.