Emotional and Cognitive Adaptation in Real Environments

Sandra Clara Gadanho*
Institute of Systems and Robotics
Torre Norte, IST, Av. Rovisco Pais 1,
1049-001 Lisbon, Portugal
email: sandra@isr.ist.utl.pt

Abstract

This paper proposes an architecture, named ALEC, for learning to make decisions in real-world environments which takes into consideration two alternative adaptation capabilities: emotional and cognitive.

ALEC brings together two different systems which have independently shown good empirical results. Previous results suggest that these systems may complement each other.

Moreover, one can argue that one of these systems embodies properties usually associated with emotions and the other with cognition. The similarities of the interaction of the two systems and that of the human emotion and cognition systems are highlighted.

1 Introduction

An agent designer, or its genes in the case of natural agents, may code the agent's behavior to some extent, but flexibility is required to deal with the unpredictability and changing circumstances of the real world. Moreover, it may be difficult for an agent's designer to predict the agent's perceptions of the environment or to think in terms of its unfamiliar sensory capabilities. These are reasons why adaptation is advantageous. Nevertheless, even the learning mechanisms can be adapted to the desired agent-environment interaction. In certain scenarios, specialized learning may have an advantage over general knowledge which is much more difficult to acquire. Animals do often benefit from domain-specific learning mechanisms which have been shaped by their specific problems through evolution [Gallistel *et al.*, 1991].

In the context of decision-making in real-world environments, autonomous adaptation is a difficult challenge. The complexity and noise in perceptual information allied with the multitude of action choices can overwhelm the agent if learning is not structured in some way. The agent designer has an important role in providing adequate learning tools. To start with, the agent must give the agent a basic value mechanism to allow it to distinguish good outcomes from bad outcomes. Building-in pre-processing of sensors and pre-constructed behaviors instead of low-level motor commands can help, but is a limiting factor on what the agent can learn.

The designer can also build in alternative adaptation mechanisms specialized in different problems which is the topic of this paper.

In [Gadanho and Hallam, 2001b; Gadanho and Custódio, 2002], an emotion-based architecture was proposed which uses emotions to guide the agent's adaptation to the environment. The agent has some innate emotions that define its goals and then learns emotion associations of environment state and action pairs which determine its decisions. The agent uses a Q-learning algorithm to learn its policy while it interacts with its world. The policy is stored in neural networks which allows to limit memory usage substantially and accelerates the learning process, but can also introduce inaccuracies and does not guarantee learning convergence [Bertsekas and Tsitsiklis, 1996].

The ALEC (Asynchronous Learning by Emotion and Cognition) architecture proposed here aims at a better learning performance by augmenting the previous emotion-based architecture with a cognitive system which complements its current emotion-based adaptation capabilities with explicit rule knowledge. The different learning capabilities of the two systems and their interaction should produce a more powerful adaptation system. The cognitive system suggested is the rule-decision system of the CLARION model [Sun and Peterson, 1998a] which is described in Section 4.1.

ALEC is based on the assumption that the cognitive system can make more accurate predictions based on rules of causality while the emotion associations have less explanatory power but can make more extensive predictions and further ahead in time.

In the next section, a description of the adaptation problem to be solved by the agent is made. This is followed by a detailed description of the reference emotion-based architecture in Section 3 and the proposed modifications in Section 4. Finally, the ALEC architecture is discussed and conclusions are drawn.

*Post-doctorate sponsored by the Portuguese Foundation for Science and Technology.

2 The Adaptation Problem

The aim of the learning mechanisms presented next is to allow an agent faced with realistic world conditions to adapt on-line and autonomously to its environment. In particular, the agent should be able to cope with continuous time and space, while constrained by limited memory, time-pressure, noisy sensors and unreliable actuators. Furthermore, the agent is required to perform a task with multiple and sometimes conflicting goals which may require sequencing of actions.

Previous experiments [Gadanho and Hallam, 2001a; Gadanho and Custódio, 2002] were carried out in a realistic simulator [Michel, 1996] of a Khepera robot — a small robot with left and right wheel motors, and eight infrared sensors that allow it to detect object proximity and ambient light. The experiments evaluate the agent in a survival task that consists of maintaining adequate energy levels in a simulated maze-like environment with obstacles and energy sources which are associated with lights the agent can sense when nearby. The agent has basically three goals: to maintain its energy, avoid collisions and move around in its environment. Moreover, the extraction of energy is complicated by requiring the agent to learn sequences of behaviors and temporarily overlook the goal of avoiding obstacles in the process. The goal of maintaining energy also requires the robot to find different energy sources in order to survive. ALEC is to be tested under the same and possibly harder conditions.

3 The Emotion-based Controller

Inspired by literature on emotions, previous work has shown that reinforcement and deciding when to switch behavior[1] can be addressed successfully together by an emotion model [Gadanho and Hallam, 2001b]. The justification for the use of emotions is that, in nature, emotions are usually associated with either pleasant or unpleasant feelings that can act as reinforcement [Tomkins, 1984; Bozinovski, 1982] and frequently pointed to as a source of interruption of behavior [Sloman and Croucher, 1981; Simon, 1967].

Later the emotion model was formalized into a goal system with the purpose of establishing a clear distinction between motivations (or goals) and emotions [Gadanho and Custódio, 2002]. In this system, emotions take the form of simple evaluations or predictions of the internal state of the agent. This goal system is based on a set of homeostatic variables which it attempts to maintain within certain bounds. The idea of homeostatic values stems from neurophysiological research on emotions [Damasio, 1994; 1999] and has been modeled previously by the DARE model [Maçãs *et al.*, 2001; Sadio *et al.*, 2001].

[1]Behavior-switching may be motivated by several factors: the behavior has reached or failed to reach its goal, the behavior has become inappropriate due to changes in circumstances, the behavior needs to be rewarded or punished. The correct timing of behavior-switching can be vital [Gadanho and Hallam, 2001a].

The architecture tested so far — see Figure 1 — is composed by two major systems: the goal system and the adaptive system. The goal system evaluates the performance of the adaptive system in terms of the state of its homeostatic variables and determines when a behavior should be interrupted. The adaptive system learns which behavior to select using reinforcement-learning techniques which rely on neural-networks to store the utility values. The two systems are described in detail next sections.

There are two further simpler systems which are hand-designed: the perceptual and behavior systems. The perceptual system is responsible for processing crude perceptions into higher-level perceptions which are expected to be more useful for the agent. The behavior system transforms simple behavior instructions into motor commands, so that the agent does not have to learn its action abilities from scratch.

3.1 Goal System

In an autonomous agent, the goal system can complement a traditional reinforcement-learning adaptive system in that it determines how well the adaptive system is doing, or more specifically, the reinforcement it is entitled to at each step. In the current work the goal system is also responsible for determining when behavior switching should occur.

The goals are explicitly identified and associated with homeostatic variables. These are associated with three different states: target, recovery and danger. The state of each variable depends on its continuous value which is grouped into three qualitative categories: optimal, acceptable, deficient and dangerous. The variable remains in its target state as long as its values are optimal or acceptable, but it only returns to its target state once its values are optimal again. The danger state is associated with dangerous values and can be coupled with urgency of recovery.

To reflect the current hedonic state of the agent a well-being value was constructed from the above. This value depends primarily on the state value of the homeostatic variables. If a variable is in the target state it has a positive influence on the well-being, otherwise it has a negative influence which is proportional to its deviation from target values.

In order to have the system working correctly two other influences on well-being are also required:

State change — when a homeostatic variable changes from a state to another the well-being is influenced positively if the change is towards a better state and negatively otherwise;

Prediction of state change — when some perceptual cue predicts the state change of a homeostatic variable, the influence is similar to the above, but lower in value and varies with the accuracy of the prediction and how soon the state change is expected.

The two goal events just described were modeled after emotions, in the sense that they result from the detection of significant changes in the agent's internal state or predictions of such changes.

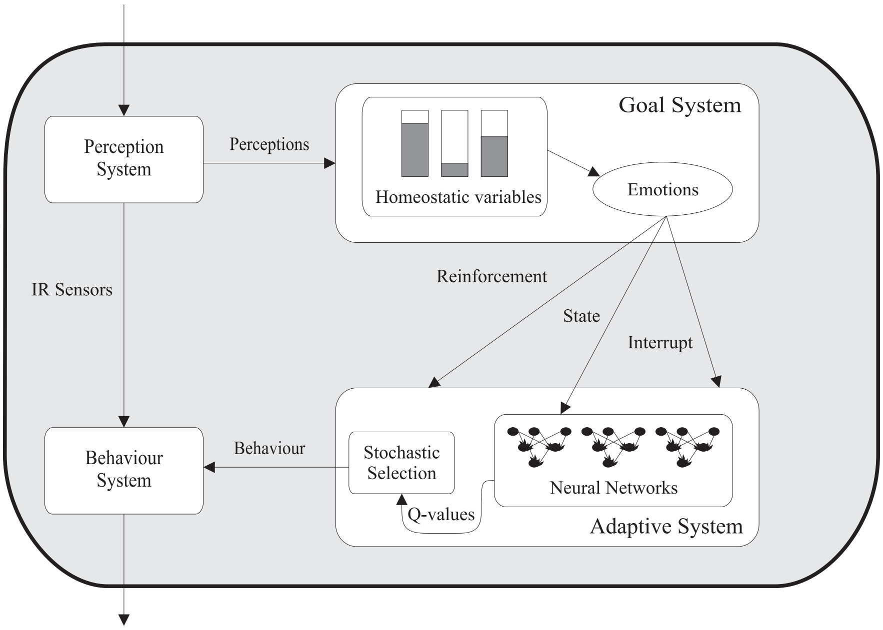

Figure 1: The emotion-based controller.

Similarly to emotions which are associated with feelings of 'pleasure' or 'suffering' depending on whether this change is for the better or not, these goal events influence the well-being value such that the information of how good the event is is conveyed to the agent through the reinforcement. One may distinguish between the emotion of happiness when a goal is achieved (or predicted to be achieved) and the emotion of sadness when of a goal state is lost (or about to be lost).

The primary influence of the homeostatic variables, on the other hand, is modeled after the natural background emotions which reflect the overall state of the agent in terms of maintaining his homeostasis [Damasio, 1999].

The goal events are also responsible for triggering the adaptive system for a new behavior selection, which is also often associated with emotions.

For the task at hand three homeostatic variables were identified: Energy, Welfare and Activity.

3.2 Adaptive System

The adaptive system implemented is a well known reinforcement-learning algorithm: Q-learning [Watkins, 1989]. Through this algorithm the agent learns iteratively by trial and error the expected discounted cumulative reinforcement that it will receive after executing an action in response to a world state, *i.e.* the utility vales (also called Q-values).

The traditional Q-learning usually uses a table, which stores the utility value of each possible action selection against every possible world state. In a real environment, the use of this table requires some arbitrary discretization of the continuous values provided by sensors. Furthermore, this can easily lead to a extremely large number of possible environment states resultant of the combination of the all the possible input values. An alternative to this method suggested by [Lin, 1992] is to use neural networks to learn by back-propagation the utility values of each action. This method has the advantages of profiting from generalization over the input space which accelerates learning and being more resistant to noise. However, neural-networks on-line training may not be very accurate.

The state information which is fed to the neural-networks is the homeostatic variable values and three perceptual values: light intensity, obstacle density and energy availability.

The developed controller tries to maximize the reinforcement received by selecting between one of three possible hand-designed behaviors:

Avoid obstacles — Turn away from the nearest obstacle and move away from it. If the sensors cannot detect any obstacle nearby, then remain still.

Seek Light — Go in the direction of the nearest light. If no light can be seen, remain still.

Wall Following — If there is no wall in sight, move forwards at full speed. Once a wall is found, follow it. This behavior by itself is not very reliable in that the robot can crash, *i.e.* become immobilized against a wall. The avoid-obstacles behavior can easily help in these situations.

At each trigger step, the agent may select between performing the behavior which has proven to be better in the past and therefore has the best utility value so far, or selecting an arbitrary behavior to improve its information about the utility of that behavior. The selection function used was based on the Boltzmann-Gibbs distribution and consists of selecting a behavior with higher probability, the higher its utility value in the current state.

4 Adding a cognitive system

This paper proposes the addition of a cognitive system to the architecture described previously (see Figure 2). The Goal System and the Adaptive System of this architecture are also referred to as the emotion system. The cognition system is expected to provide an alternative decision-making process to the emotion system. It relies on more traditional A.I. reasoning based on a collection of important discrete event instances. This alternative memory representation has two main advantages: it is not prone to the inaccuracies due to neural-network over-generalization; and it allows the use of more conventional A.I. techniques such as planning.

The cognitive system should collect information independently and step in to correct the emotion system's decisions. The cognitive system proposed is the rule-based system of the CLARION model which is described next.

4.1 The CLARION model

The CLARION model [Sun and Peterson, 1998a; Sun *et al.*, 2001] is a hybrid cognitive model which addresses the problem of bottom-up on-line learning of low-level skills and high-level declarative knowledge.

It consists of two decision-making layers, each with different adaptation capabilities. The bottom-layer is a Q-learning system using neural-networks which is very similar to the adaptive system described in Section 3.2. The top-layer is a rule-based system which is distinguishable from other rule systems in that it is not derived of an a-priori pre-constructed set of rules given externally. Instead, rules are extracted from the agent-environment interaction experience through the mediation of low-level skills [Sun and Peterson, 1998b]. Other models are usually top-down, *i.e.* through practice the agents turn high-level knowledge into usable procedural skills [Sun *et al.*, 2001]. Nevertheless, a-priori knowledge can still be easily given to the system in the form of rules and if these are useful they will actually be assimilated into procedural knowledge by the system [Sun *et al.*, 2001].

Each individual rule is triggered by specific environmental conditions and suggests an action choice. Rule acquisition and revision is based on gradual accumulation of statistics, but is done in a one-shot and all-or-nothing fashion. If some action is found successful then the agent extracts a rule correspondent to the decision made and adds it to its rule set. Subsequently, the agent verifies the usefulness of the rule by applying it: if the outcome is successful the agent tries to generalize it by making it cover more environmental states, otherwise it will make it more specific and exclusive of the current case (it may even delete it).

The success of the agent is measured in terms of its immediate reinforcement and in terms of the difference of Q-values between the state where the decision was made and the state reached after the decision was taken. This means that rule learning takes into consideration the information collected by the bottom-level. Rule learning is limited to those cases for which the model has sufficient experience and leaves the other cases to the bottom-level which makes use of its generalization abilities [Sun and Peterson, 1998a].

The action decision taken at each moment may rely on a top-level or a bottom-level suggestion. If the top-level has a suggestion then the suggestion to be used is selected probabilistic based on the recent relative competence of the top-level and the bottom-level [Sun and Peterson, 1998a]. This means that as the top-level becomes more competent it is used more often.

The authors report a synergy between the two levels [Sun and Peterson, 1998a] and attribute it to the complementary representations (discrete *vs.*continuous) and learning methods (one-shot rule-learning *vs.*gradual Q-value approximation) of the two levels.

On the one hand, the top-level cannot learn without the bottom-level, since it has no form of temporal credit assignment and it needs the bottom-level's long-term predictions. On the other hand, the bottom-level preforms worse without the help of the top-level due to the inaccuracies of the back-propagation networks (*i.e.* the blurring effect of their generalization abilities, which can be partially alleviated when the crisp top-level is added) [Sun *et al.*, 2001]. Rules complement the function approximator by detecting and correcting over-generalization [Sun and Peterson, 1998a].

5 Discussion

ALEC approach implies that while emotion associations may be more powerful in its range capabilities, they lack explanation power and may introduce errors of over-generalization.

Cognitive knowledge on the other hand is restricted to learning about simple short-term relations of causality. Its information is more accurate, but at a price. Since it's not possible to store and consult all the single events the agent experiences, it selects only a few instances which seem most important.

In summary, the two learning capabilities solve the problem of too much information provided by the agent-environment interaction in two different ways: one stores all events, but no information to distinguish between individual events, all events are mixed together; the other one only extracts the most significant events.

The way the emotion level influences the cognitive level is akin to Damásio's somatic-marker hypothesis [Damasio, 1994]. In his hypothesis, Damásio suggested that humans associate high-level cognitive decisions with special feelings which have good or bad connotations dependent on whether choices have been emotionally associated with positive or negative long-term outcomes. If these feelings are strong enough, a choice may be immediately followed or discarded. Interestingly, these markers do not have explanation power and the reason for the selection may not be clear. In fact, although the decision may be reached

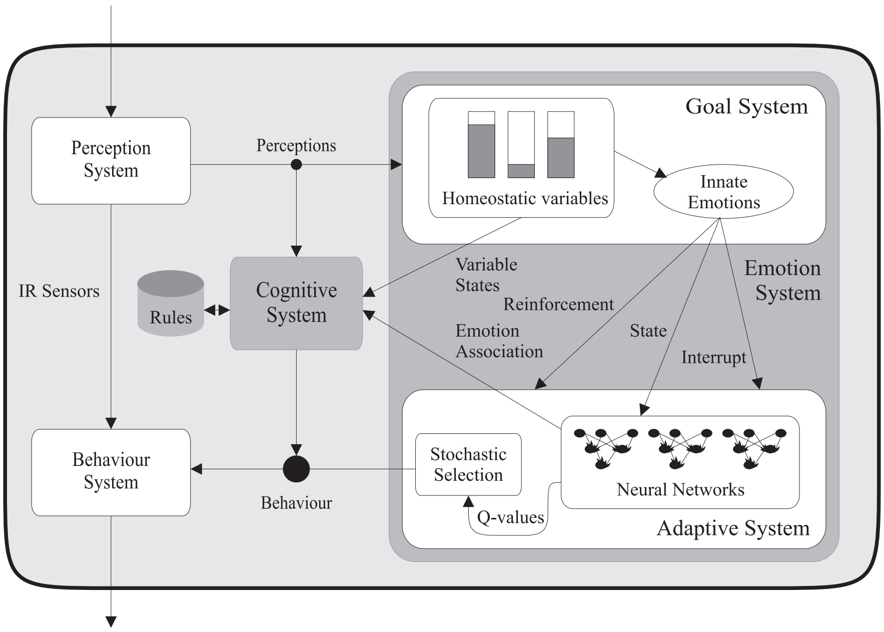

Figure 2: The ALEC architecture.

easily and immediately, the person may feel the need to subsequently use high-level reasoning capabilities to find a reason for the choice. Meanwhile, a fast emotion-based decision could be reached which depending of the urgency of the situation may be vital.

ALEC shows similar properties, when it uses emotion associations to guide the cognitive system. Furthermore, the rule system can correct the emotion system when this reaches incorrect conclusions. Knowing the exceptions from previous experiences, it may choose to ignore the emotion reactions, which although powerful can be more unreliable.

The ALEC architecture is akin to the full CLARION model, but has an important difference: the existence of a well-defined goal system in ALEC. This opens extra possibilities for the development of the rule system. For instance, the rule system can be extended to treat separately the various goals of the system and learn how to individually reach the target states of each one of the homeostatic variables. In fact, the rule system can specialize in learning about transitions in the agent's internal state.

A related approach is the DARE model [Ventura and Pinto-Ferreira, 1999; Maçãs *et al.*, 2001; Sadio *et al.*, 2001] which is particularly concerned with the dual evaluation of the perceptual stimulus [Damasio, 1994; LeDoux, 1998]. In this model, there are two paths for stimulus evaluation: the perceptual and the cognitive. The perceptual designates the "quick and dirty" processing usually associated with emotions. The cognitive attempts to be a more sophisticated evaluation provided by higher-level reasoning. These layers may have their own separate learning mechanisms for adapting their evaluations, but in the experiments the perceptual evaluation is often implemented as innate fixed knowledge and the cognitive layer always learns from scratch.

In the DARE model, the perceptual layer extracts relevant features and the cognitive layer task is to identify objects. Nevertheless, a recurrent feature of the implementations which shares with ALEC, is that the perceptual level has a non-differentiated evaluation of events by their main characteristics while the cognitive level accumulates a set of individual instances of events.

A further advantage of ALEC is the infrastructure for endowing the agent with innate knowledge about the world in two distinct forms, as preferences/dislikes at the emotion system or as simple action rules at the cognitive system. For different problems of the same task, the knowledge may be more evident to the designer one way or the other.

6 Conclusion

In ALEC, extra designer knowledge is put into the complexity of the learning system which is based on two flexible structures endowed with different learning capabilities. These structures which are modeled after the human emotional and cognitive reasoning abilities give the agent a more powerful adaptation capacity than simpler learning mechanisms.

The existence of the cognition and emotion as two interacting systems, both with important roles in decision-making has been recently advocated by neuro-physiological research [LeDoux, 1998; Damasio, 1994]. DARE is a prescriptive model of Damásio's ideas which makes use of this concept of dual decision path. Although its implementations thus far follow simpler computational approaches with emphasis on different theoretical aspects, the basic theoretical ideas are in tune with the ones presented here.

ALEC has already been partially tested. Extensive empirical results on the performance of the emotion-based architecture have been presented elsewhere [Gadanho and Hallam, 2001b]. The results demonstrated that it was quite competent when compared with more traditional approaches, in spite of the limited capabilities of the knowledge representation. The rule system suggested for the cognitive system has also been previously tested, with positive experimental results. Furthermore, these results suggest that the introduction of this extra system may complement the capabilities of the emotion system. Therefore, it is expected that the proposed modifications to the architecture will enhance the learning performance of the agent.

Hopefully, ALEC will be an example of complex emotion and cognitive systems with very different adaptation and decision-making capabilities which successfully cooperate in the control of an agent faced with real world problems.

References

[Bertsekas and Tsitsiklis, 1996] D. Bertsekas and J. Tsitsiklis. *Neuro-Dynamic Programming.* Athena Scientific, Belmont, MA, 1996.

[Bozinovski, 1982] Stevo Bozinovski. A self-learning system using secondary reinforcement. In R. Trappl, editor, *Cybernetics and Systems*, pages 397–402. Elsevier Science Publishers, North Holland, 1982.

[Damasio, 1994] Antonio R. Damasio. *Descartes' error — Emotion, reason and human brain.* Picador, London, 1994.

[Damasio, 1999] Antonio Damasio. *The feeling of what happens.* Harcout Brace & Company, New York, 1999.

[Gadanho and Custódio, 2002] Sandra Clara Gadanho and Luis Custódio. Learning behavior-selection in a multi-goal robot task. In *NAISO ICAIS proceedings*, 2002. *In Press.*

[Gadanho and Hallam, 2001a] Sandra Clara Gadanho and John Hallam. Emotion-triggered learning in autonomous robot control. In Dolores Cañamero and Paolo Petta, editors, *Grounding emotions in adaptive systems — A Special Issue of Cybernetics and Systems*, volume 32, pages 531–559. Taylor & Francis, July 2001.

[Gadanho and Hallam, 2001b] Sandra Clara Gadanho and John Hallam. Robot learning driven by emotions. *Adaptive Behavior*, 9:1, 2001. *In Press.*

[Gallistel *et al.*, 1991] C. R. Gallistel, Ann L. Brown, Susan Carey, Rochel Gelman, and Frank C. Keil. Lessons from animal learning for the study of cognitive development. In Susan Carey and Rochel Gelman, editors, *The Epigenesis of Mind: Essays on biology and Cognition.* L. Erlbaum Associates, 1991.

[LeDoux, 1998] Joseph E. LeDoux. *The Emotional Brain.* Phoenix, London, 1998.

[Lin, 1992] Long-Ji Lin. Self-improving reactive agents based on reinforcement learning planning and teaching. *Machine Learning*, 8:293–321, 1992.

[Maçãs *et al.*, 2001] Márcia Maçãs, Paulo Couto, Carlos Pinto-Ferreira, Luis Custódio, and Rodrigo Ventura. Experiments with an emotion-based agent using the DARE architecture. In *Proceedings of the AISB'01 Symposium on Emotion, Cognition and Affective Computing*, pages 105–112, University of York, U. K., March 2001.

[Michel, 1996] Olivier Michel. *Khepera Simulator* package version 2.0: Freeware mobile robot simulator written at the University of Nice Sophia–Antipolis, March 1996. Downloadable from the World Wide Web at `http://wwwi3s.unice.fr/~om/khep-sim.html`.

[Sadio *et al.*, 2001] Rui Sadio, Gonçalo Tavares, Rodrigo Ventura, and Luis Custódio. An emotion-based agent architecture application with real robots. In Lola Cañmero, editor, *AAAI Fall Symposium on Emotional and Intelligent II: The tangled knot of social cognition*, pages 117–122. Menlo Park, California: AAAI Press, 2001. Technical report FS-01-02.

[Simon, 1967] H. A. Simon. Motivational and emotional controls of cognition. *Psychological Review*, 74:29–39, 1967.

[Sloman and Croucher, 1981] Aaron Sloman and Monica Croucher. Why robots will have emotions. In *IJCAI'81 — Proceedings of the Seventh International Joint Conference on Artificial Intelligence*, pages 2369–71, 1981. Cognitive Science Research Paper 176, Sussex University.

[Sun and Peterson, 1998a] R. Sun and T. Peterson. Autonomous learning of sequential tasks: experiments and analysis. *IEEE Transactions on Neural Networks*, 9(6):1217–1234, November 1998.

[Sun and Peterson, 1998b] R. Sun and T. Peterson. Some experiments with a hybrid model for learning sequential decision making. *Information Sciences*, 111:83–107, 1998.

[Sun *et al.*, 2001] R. Sun, E. Merrill, and T. Peterson. From implicit skills to explicit knowledge: a bottom-up model of skill learning. *Cognitive Science*, 25(2):203–244, 2001.

[Tomkins, 1984] Silvan S. Tomkins. Affect theory. In Klaus R. Scherer and Paul Ekman, editors, *Approaches to Emotion.* Lawrence Erlbaum, London, 1984.

[Ventura and Pinto-Ferreira, 1999] Rodrigo Ventura and Carlos Pinto-Ferreira. Emotion-based agents: Three approaches to implementation (preliminary report). In Juan D. Velsquez, editor, *Workshop on Emotion-Based Agent Architectures*, Seattle, U. S. A., 1999. Workshop of the Third International Conference on Autonomous Agents.

[Watkins, 1989] C. Watkins. *Learning from delayed rewards.* PhD thesis, King's College, Cambridge, 1989.