The Evolution of Simple Affective States in Multi-Agent Environments

Matthias Scheutz
Department of Computer Science and Engineering
University of Notre Dame
Notre Dame, IN 46556, USA
mscheutz@cse.nd.edu

Abstract

We propose a research strategy to study the evolution of affective states and analyze the requirements for simulated environments to be appropriate for experiments with affective agent architectures. We present the simulation model and agent architecture used in our experiments to demonstrate that (1) primitive emotional states (such as "fear" and "anger") and primitive motivational states (such as "hunger" and "thirst") can play an important role in the control and coordination of agents in agent societies, and (2) such states are very likely to evolve (in certain environments).

Introduction

The potential of affective states as efficient and powerful coordinators and controllers of agent behavior has become appreciated in recent years, as witnessed by the increasing number of research projects on this topic (e.g., Maes (1991), Tyrrell (1993), the LEE model (Menczer and Belew 1996), Spier and McFarland (1998), the Cathexis model (Velásquez 1997), the Abbot model (Cañamero 1997) and its extensions, the Kismet model (Breazeal 1998), Seth (2000), the various models by the SAB community (e.g., see Meyer et al. 2000) and many others.[1] Especially in complex and unpredictable environments where agents have limited resources (e.g., computational power, memory capacity, etc.) and sensory information is not reliable, classical (rationality-based) decision methods to determine the best action for an agent (from its current sensory information, its internal state, its current goals, its knowledge, etc.) are not applicable. In such circumstances, where it is impossible to provide complete, perfect, and reliable information, mechanisms relying on affective states (such as motivations, desires, attitudes, preferences, moods, and some emotions) can be very effective and can serve as context-sensitive initiators modulators, and regulators of an agent's behavior. While natural systems are the canonical models for affect-based control systems, little is known about the evolutionary trajectories of affective mechanisms, i.e., under what conditions various kinds of affective states are likely to evolve in competitive multi-species environments and what the evolutionary advantages of affective states are *qua* control states over other control states. An answer to these questions may help us understand how affect is grounded in the interaction of agents with each other and their environments.

In this paper, we attempt to contribute to answering these questions by focussing on two different kinds of affective states: primitive "motivational states" (or *drives*) and primitive "emotional states". For easy reference, we will put the labels "hunger" and "thirst" on the former, and "fear" and "anger" on the latter, while keeping in mind that the states used in the simulations below may bear varying resemblance to the various states with the same labels found in nature. We first sketch a research strategy to study affective states in agent architectures, and then demonstrate this strategy by examining the evolutionary trajectory of particular agents without the above affective states to agent with those states in a simulated environment. While we stress the functional roles and evolutionary advantages provided by many affective mechanisms, we do not claim that *all* affective states are biologically advantageous (some may be by-products of useful mechanisms without being useful in themselves, e.g., see Sloman 2000a). An analysis of disadvantageous affective processes, however, is beyond the scope of this paper.

A Strategy for Studying Affective States and their Origin

The notion of an "affective state" state is a so-called cluster concept, which defies the usual characterization of classical concepts in terms of necessary and sufficient conditions. Worse yet, most of its subspecies are cluster concepts themselves, in particular the notion of "emotion" (there are numerous different, partly incompatible characterizations of what emotions are in psychology alone, e.g., see Griffiths 1997). Not surprisingly, neither terminology, nor conceptual underpinnings of different forms of affect form a uniform picture in AI either. While some see emotions as special kinds of motivations (e.g., Breazeal, 1998), others draw a distinction between motivations and emotions (e.g., Cañamero 1997). We believe that these discrepancies result to a large extent from the fact that mental concepts seem to be intrinsically *architecture-based* concepts. Hence, a satisfactory analysis of such concepts will need to show how affective states depend on and contribute to important capa-

[1] For space reasons we only list projects here that are also referred to later in this paper. For an overview of other models, see for example Pfeiffer (1988) or Picard (1997).

bilities within an agent architecture (Sloman 2000a).

Without being able to go into any detail in this paper, we suggest that what underwrites the common use of the term "affective" is essentially the concept of a control state (Sloman 1993), and that while not every control state is an affective state, the subclass of affective control states can be characterized by a distinction between "belief-like" and "desire-like" states: if a system's behavior changes an internal state to make it fit reality, then this state is "belief-like", whereas if the system changes reality to make it fit the internal state, then it is "desire-like" (Scheutz and Sloman 2001). It is this distinction between belief-like and desire-like control states that can give us a handle on how to construe affective states, namely as *desire-like control states*, whose role is initiating, evaluating and regulating, internal or external behavior (as opposed to merely acquiring, interpreting, manipulating, or storing information that might or might not be used in connection with affective states to initiate or control behavior). We shall use the term "affective states" in this sense for the remainder of this paper.

Since affective states–the way we construe them–are the springs and guides of action (and sometimes disruptive side-effects of such springs and guides) in natural systems, we would like to understand the logical space of possible affective states to be able to utilize them in artificial systems. There seem to be two partly overlapping classes of three questions each that are relevant in the context of our understanding of affective states. The first class concerns affective states as they occur in nature, asking (1) what affective states are and what different kinds of affective states there are, (2) how and why affective mechanisms came about, and (3) what their function (if they have a function) is in information processing architectures. The second class asks similar questions about a wider set of affective states in actual organisms, theoretically possible biological organisms and artificial agents. This leads to an additional question: (4) how can such affective mechanisms be incorporated in agent architectures and implemented in real and synthetic agents?

Conceptual analyses of affective states are mostly targeted at answering questions (1) and (3), investigations in the empirical sciences mostly attempt to answer questions (2) and (3) (especially in the last decade cognitive scientists paid increasing attention to the evolutionary context, in which affective states have developed). Successful implementations of AI models, on the other hand, which employ (simple) affective states to control the behavior of simulated or real agents, provide (partial) answers to questions (3) and (4) (with respect to the implemented model), but do not answer questions (1) or (2) (for one, because models that do not contrast their implementation with alternative ways of achieving the same goal, are in a sense mere "existence proofs" showing that certain affective states can assume a particular functional role in a particular system).

We believe that an answer to these question will likely not come forth from independent inquiries, but from the interplay of conceptual analyses, empirical findings and concrete experiments with agent architectures. The proposed research strategy then is to start with a notion of affective state, which is applicable to natural systems, determine/define its function in a particular agent architecture and subsequently try to explore the properties of this state for concrete agents in different environments with the goal of extending the notion to more complex cases. This includes investigating ways in which slight changes in environments can change the tradeoffs between design options for the architecture and hence for the functional role of the affective state. Such explorations of "neighborhoods in design and niche space" (e.g., Sloman 2000b) will help us understand what the competitive advantage of a particular change in architecture or mechanism might be in a particular environment, and how the benefits change in slightly different environments.

Requirements for the Experimental Setup

To be able to study the origins of affective states from an evolutionary perspective and effectively experiment with different kinds of agent architectures, a genuine *artificial life simulation environment* is required, within which different species of agents (with different architectures and possibly different bodies) can *coexist* and *procreate*. Both requirements are crucial; the first, because affective states in natural systems did not evolve in isolation, but rather in competing multi-species societies. Hence, to fully appreciate the benefits of affective states, we need to study the tradeoffs between different control architectures in competition with each other. A model employing affective states in the control of a particular isolated agent or a group of agents *with identical architectures* is necessarily silent about the evolutionary advantage of affective control over other ways of controlling and regulating behavior (e.g., by virtue of various kinds of non-affective reactive or deliberative processes) in a multi-species environment. The fact that agents of one kind perform better than agents of another kind if tested independently does not shed any light on their performance in mixed groups.

The second requirement is equally important, because classic genetic algorithms (GAs) assess the fitness of agents based on a static, predetermined fitness function and can hardly (if at all) do justice to the dynamics of the local interactions of agents with their (changing) environments, which in the end determines reproductive success (e.g., see Kauffman 1995). There are several problems with specifying fitness explicitly besides evolutionary plausibility. For one, it is not clear what architectural features to select for if the task at hand is to evaluate the role and potential of affective states in different agent architectures from an evolutionary perspective. Furthermore, as agents and their architectures change over time together with the environment, adaptive fitness changes as well, which would have to be somehow reflected in the fitness function (for a more detailed description of the differences between *exogenous* and *endogenous* fitness and some reasons why endogenous fitness is to be preferred in such a simulation setup, see Menczer and Belew 1996). In general, it seems that we should refrain from imposing any particular behavioral criteria on agents other than their ability to procreate so as to not bias their evolutionary trajectories.

This is not to say that GAs cannot be employed successfully to evolve functioning agents with certain kinds of affective states. In fact, our results below indicate that it should be (relatively) easy to evolve agents with controllers that implement certain primitive affective states with a classic GA, if they evolve even in competitive multi-agent environments. What simply does not follow automatically from classic GA experiments is that the same results could have been obtained if fitness had been assessed implicitly by allowing the agents to procreate in competition with other species or subspecies (unless all the factors that could possibly lead to and be responsible for the procreation of an agent are part of the explicit fitness function).

Other desiderata include spatial continuity (to eliminate any potential influence of grid structures), temporal sensitivity (to be able to study temporal trade-offs of actions and processing mechanisms), at least two resources that agents need to obtain (to make the decision problem interesting, e.g., Tyrrell 1993, or Spier and McFarland 1998), and Lamarckian mutation mechanisms (to be able to control modifications and extensions of certain components of an architecture).[2]

We have developed the SimWorld[3] model based on the above requirements in order to be able to study the origins and roles of affective states in agent societies with possibly many different kinds of agents. In the following, we will first describe the experimental setup, the agents and their architectures used in the experiments, and then present the main results.

The SimWorld Simulation Environment

SimWorld consists of an unlimited continuous surface populated with various spatially extended objects such as various kinds of agents, static obstacles of varying size, and food and water sources, which pop up within a particular area (usually of about 700 by 700 units) and disappear after a pre-determined period of time, if not consumed by agents earlier. Agents are in constant need of food and water as moving consumes energy and water proportional to their speed–even if they do not move, they will still consume a certain amount of both. When the energy/water level of an agent drops below a certain threshold ω, agents "die" and are removed from the simulation. They also die and are removed, if they run into other agents or obstacles.

All agents are equipped with exteroceptive "sonar", "smell", and "touch" sensors. Sonar is used to detect obstacles and other agents, smell to detect food and water, and touch to detect impending collisions with agents or obstacles as well as consumable food and water sources. In addition, the touch sensor is connected to a global alarm system, which triggers a reflex beyond the agent's control to move the agent away from other agents and obstacles. These movements are somewhat erratic and will slightly reorient the agent (thus helping it to get out of "local minima"). Furthermore, agents have two proprioceptive sensors to measure their energy and water levels, respectively.

On the effector side, they have motors for locomotion (forward and backward), motors for turning (left and right in degrees) and a mechanism for consuming food and water (which can only be active, when the agent is not moving). When agents come to a halt on top of a food or water source, their ingestion mechanism suppresses the motors for locomotion until the item is consumed, which will take a time proportional to the amount of energy/water stored in the food/water source depending the maximum amount of food/water an agent can take in at any given time.

After a certain age α (measured in terms of simulation cycles), agents reach maturity and can procreate asexually. Since the energy for creating offspring is subtracted from the parent, agents will have a variable number of offspring depending on their current energy level (from 0 to 4), which pop up in the vicinity of the agent one at a time. Since a mutation mechanism modifies with a certain probability μ some of the agent's architectural parameters (e.g., such as connection weights in a neural network), some offspring will start out with the modified parameters instead of being exact copies of the parent. Note that both parameters, α and ω, can be used to specify, whether the simulation is used as an exogenous or as an endogenous fitness model.

Agents, Architectures and Behaviors

While different agents may have different (implicit) short-term goals at any given time (e.g., getting around obstacles, consuming food, reaching a water source faster than another agent, or having offspring), common to all of them are two (implicit) long-term goals: (1) *survival* (to get enough food/water and avoid running into obstacles or other agents), and (2) *procreation* (to live long enough to have offspring).

In the following experiments, we study different kinds of related agents, which all possess the same architectural components (but not all the same links among them). All agents process sensory information and produce behavioral responses using a schema-based approach (Arkin 1989). Let $Ent = \{f, w, o, a\}$ be an index set of the four types of objects *food*, *water*, *obstacle*, and *agent*–all subscript variables will range over this set unless stated otherwise. For each object type in Ent, a force vector F_i is computed, which is the sum, scaled by $1/|v|^2$, of all vectors v from the agent to the objects of type i within the respective sensory range, where '$|v|$' is the length of vector v. These four *perceptual schemas* are then mapped into motor space by the transformation function $T(x) = g_f \cdot F_f + g_w \cdot F_w + g_o \cdot F_o + g_a \cdot F_a$ for $i \in Ent$, where each g_i is the respective gain value. These gain values are provided by the output layer of a three-layer *interactive activation and competition* (IAC) neural network with four input units in, four hidden units hid, and four output units out (Rumelhart and McClelland, 1986) via individual scaling functions $f_i(x) = x \cdot c_i + b_i$ (where b_i is

[2] Note that this is for methodological reasons only. As long as these mutation operations are feasible using Darwinian mutation, we can justify performing operations directly on the architecture instead of performing them on genetic representations. As an aside, it is always possible to regard architectures as representations of themselves, although it is doubtful that organisms would use such an uncompressed code.

[3] The SimWorld environment builds on the SimAgent toolkit developed by Aaron Sloman and colleagues at the University of Birmingham, which is freely available at http://www.cs.bham.ac.uk/research/simagent/.

the *base gain value* and c_i the scaling factor for the activation of out_i). The input layer is connected (again via similar scaling functions) to the internal water (in_w) and energy level sensors (in_f) as well as the global alarm mechanism (which sends an impulse to in_o or in_a units depending on whether the alarm was triggered by an impending collision with an agent or an obstacle). Note that neural networks employed in other simulations to control the behavior of agents (Menczer and Belew 1996, Seth 2000, et al.) usually compute the mapping from sensors to effectors, while the neural network here is intended to implement the affective system, thus adding another layer on top of the input-output mapping (which is accomplished in a schema-based manner; of course, this mapping, in turn, could have been implemented as neural network as well).

The choice of IAC units over standard perceptrons is based on their update rule, which is particularly suited to implement important temporal features of affective states in that it (1) takes into account the *previous activation* (hence, can be used to implement "inner states"), and (2) incorporates a *decay term* to raise or lower the activation to a predetermined *base level* (both features that seem to be typical of the temporal development of certain affective states, e.g., basic emotional states). Very similar update rules (with only minor differences to IAC units) are also used in other implementations of systems with affective states, although they usually go by a different name (e.g., in the Cathexis or Kismet models).

Although fully connected IAC networks are possible, we will focus on a subset of networks at this point to avoid complexity, where weights between in_i and hid_i are always non-zero and weights between hid_i and out_i, call them ow_i, may be non-zero, all other weights being zero. In *basic agents*, then, each ow_i is zero and as a result the corresponding gain value $g_i = b_i$, i.e., constant. Consequently, the behavior of such agents is completely determined by their inputs: inner states, as possibly implemented by the hidden units, *do not contribute* to their behavior, which is entirely reactive. Basic agents are contrasted with *extended agents*, where some ow_i are non-zero and gain values in T can consequently vary depending on the state of the neural network.

As one might expect, the differences in behavior between the various kinds of agents can be very subtle as the influence of the hidden units on the gain values can be very gradual, and hence very difficult to detect. It is therefore crucial to look at a time-frame larger than the life-time of a single agent to be able to evaluate the advantages and disadvantages of different weight values, in particular, in competitive multi-agent environments. In fact, most tradeoffs are only visible in simulations of many generations of agents in different combinations under different environmental conditions. Nevertheless, it is possible to sketch a few general behavior tendencies. The basic agents, for example, always behave in the same way given that their gain values are constant: with positive $g_f = g_w$ they behave like the "consume nearest" strategy in environments without obstacles (Spier and McFarland 1998). Negative $g_o = g_a$ values will make them avoid obstacles and other agents. In extended agents (with the same gain values) the degree to which they engage in the respective behaviors will in addition to the sign and strength of the weights depend on the activation of the respective hidden units and hence vary from time to time (e.g., they tend to avoid food, if they are not "hungry").

The Evolution of Simple Emotional States

We have shown elsewhere (although in a slightly different setup, see Scheutz and Sloman 2001) that agents with positive ow_f and ow_w weights, call them *motivational agents*, are likely to evolve from basic agents independent of many environmental conditions such as the frequency of appearance of new food and water sources, or the numbers and initial distributions of food and water sources, obstacles and agents. We argued that these agents implement two primitive motivational states, i.e., "hunger" and "thirst" drives. Here we extend these results to agents with additional affective states, the primitive emotional "fear" or "anger" state (we first present the results and then justify the labels in the next section).

We start with environments populated only by motivational agents and allow for mutation of w_o and w_a by the fixed mutation factor $\tau = 0.05$. Whenever an agent has offspring, the probability μ for modification of any of the two weights is 1/3 (i.e., 1/6 for increase or decrease by τ, respectively). The results are shown in Table 1: in 8 out of 20 runs of the simulation, where 20 agents were randomly placed in an environment with 30 obstacles, some (mutated) agents survived after the maximum number of 100000 update cycles (which is the equivalent of 400 to 500 generations given that the average life-time of agents is around 220 cycles in those simulations). Table 2 shows average and standard distribution of the various weight values that were evolved by each surviving group. Note that surviving groups are extremely uniform, i.e., agents within such groups all have very similar weights. If we correlate the number of surviving agents (abbreviated by 's') with the magnitudes of their respective weights, then we find a strong anti-correlation of -0.97 between s and positive w_o (indicating that being attracted by obstacles is not conducive to survival), little correlation between s and positive w_a (indicating that being attracted by other agents may only do very little for survival), but quite strong correlations of 0.66 between s and negative w_o, and 0.79 between s and w_o (indicating that being repelled by obstacles and especially other agents will facilitate survival). We also computed various other correlations between the two weights and groups of agents (e.g., taken over the whole course of evolution or over a restricted period) and have found a similar picture with respect to the ordering of the correlations (although with different values).

Analysis

Given the above results, how can we say that agents do or do not implement certain affective states? First, it is crucial to distinguish between (at least) two classes of affective states that are supported by the architecture and grounded in a difference between the inputs to the neural net and their connection to entities in the world: inputs coming from the energy and water level sensors can be regarded as indicat-

Table 1: The result of placing 20 motivational agents with $w_h = 0.7$ and $w_t = 0.5$ in an environment with 30 obstacle using a plant rate of 0.25 and water rate of 0.25 averaged over 20 runs of 100000 simulation cycles each.

	μ	σ	Con
Alive	4.85	7.37	2.85
Thirst	157.25	103.87	40.16
Hunger	1011.25	631.39	244.11
Crashed	2064.85	1299.83	502.54

Table 2: The number of surviving agents and the average values of their evolved w_o and w_a weights for the 8 simulations with any surviving agent.

Num.	w_o		w_a		Kinds	
	μ	σ	μ	σ	Obsts	Agents
4	0.0	0.07	-0.41	0.03	-	f
4	0.48	0.05	0.31	0.06	a	-
7	0.0	0.04	-0.01	0.13	-	-
12	-0.43	0.08	0.68	0.08	f	a
13	-0.09	0.05	0.16	0.03	-	-
16	0.24	0.12	0.51	0.10	-	a
18	0.11	0.11	-0.61	0.09	-	f
23	-0.56	0.07	-0.79	0.08	f	f

ing discrepancy values between the actual level and the normal/optimal level of a controlled "physiological" variable. Hence, the links from proprioceptive sensors to input units, to hidden units, to output units, and finally to gain values can be seen to implement the processing of an *error signal*, which indicates that a homeostatic value is outside of its normal range, to adjust behavior. Such processes are typically identified with *drives* (e.g., see the various references to McFarland's earlier works in Spier 1998). In the above case, by virtue of the connections to the (simulated) energy and water level sensors, these correspond to "hunger" and "thirst" drives.

The other two links, based on inputs coming from the global alarm, however, do not seem to implement drives. For one, they are not connected to a proprioceptive sensor that measures the state of an internal variable. Rather, they are connected to a mechanisms that *can be used* to measure the frequency of encounters with certain kinds of objects over a particular period of time. While one alarm triggering might not have much effect at all, high frequencies of alarm triggering will lead to high activations of the corresponding hidden unit, which in turn exert influence on the associated gain value in T. This influence can be seen as an *amplifying* or *diminishing* modification of the behavior as determined by the drives, which is typical of (some construals of) emotional states (e.g., see Cañamero 1997 for a similar view). More specifically, the implemented states seem to correspond to so-called "primary emotions" (e.g., Sloman 2000a) in that they (1) play a regulatory role, (2) are engaged automatically (by virtue of the global alarm system), and (3) alter the internal state of the agent and consequently its behavior. Note that the relation between sensor activation and hidden unit activation is not as direct as in the case of drives, but rather indirect involving integration over time. Furthermore, the intensity level of these emotional states will return to normal by itself by virtue of the decay rate of the hidden units unless new interfering alarm triggerings keep it up, in contrast to the activations of the drive states, which are tightly coupled to the activations of the proprioceptive sensors.

The process of building up activation over time, which is not directly related to the activation level of some internal sensor, but to the frequency of external simulation, seems to be typical of emotional states like "fear" and "anger". We suggest that depending on the signs and strengths of their w_o and w_a weights, agents will implement one of the two states: with a sufficiently strong negative weight, which creates a repulsive force causing the agent to avoid either other agents or obstacles, a "fear-like" state will be implemented, whereas with a sufficiently strong positive weight, which creates an attractive force leading to increasing insistence on the agent's part to continue its current movement–a behavior that could be described as aggressive–an "anger-like" state will be implemented. "Sufficiently strong" in this context means "to be able to influence the behavior significantly", which is usually the case for absolute weight values greater than about 0.5 (+/- 0.1), a level reached by half of the weights in surviving agents (the results are summarized in the rightmost column of Table 2, where 'a' stands for "anger", 'f' for "fear", and '-' for "no state"). However, a word of caution seems appropriate at this point as we are aware that attributions of affective states to agents of the above kind, which depend on whether a variable has value greater than a given threshold, are highly problematic and it may be better to speak of *degrees of affective influence* in such circumstances (in particular, if we are looking at intermediary stages of evolutionary trajectories).

Discussion and Future Work

The above experiments demonstrate the research strategy suggested earlier, which we believe will help us understand the role and origins of affective states as well as the potential uses affective states can be put to in the control of agents. Furthermore, the experiments confirm that if there are architectural components that can implement them, affective states like "hunger", "thirst", "fear", and "anger" are likely to evolve, even in very competitive multi-agent environments. The degree of competitiveness of these environments is apparent from the fact that on average any basic agent is still alive after 100000 update cycles in only 1 out 20 runs of a simulation *without mutation*. This goes to show that the evolved affective states are not only beneficial to the individual agent, but also lead to behavior, which benefits the whole species. More specifically, agents use an improved version of the *cue xcilte f* strategy (e.g., Spier and McFarland 1998) to forage for food and water, which takes the "clumpiness" (Seth 2000), i.e., the degree to which agents tend to stick together, into account.

We used a schema-based agent architecture (quite common in behavior-based robotics, but rather unusual for such an evolutionary setting) to show how affective states can be implemented in components linking proprioceptive sensory inputs and internal global alarm mechanisms to components implementing the gain values of motor schemas. The causal linkages effected by this architecture, which enable affective states to exert influence on the agent's behavior at any given time, is what makes them *affective states* in the first place. Furthermore, the architecture obviates the need for explicit action-selection mechanisms and explicit representations of behaviors at the architecture level, which we believe to rest on a conflation of a behavioral and a mechanistic level of description and explanation (see also Seth 2000). In other words, our agents can still can be engaged in a "go-towards-food" behavior, then get interrupted by a "veer-around-obstacle" behavior, become attracted to water and engage in a "deviate-from-original-course-to-drink-water" behavior, and so on without the need for similarly labeled, functional components in the agent architecture (as seems to be very common, e.g., Maes 1991, Velásquez 1997, Breazeal 1998, et al.). A unwanted consequence of such explicit representations of behavior is that affective states (e.g., "hunger") are often, in our view unnecessarily, *associated* with a particular behavior (e.g., "seeking-food") at the architecture level. Such design decisions, however, need to be justified and a case needs to be made that these states are indeed "affective states" and not merely local parameters that exert influence on the behavior "only" when the behavior is "active" or when the behavior is selected (e.g., by an implicit action-selection mechanism using a "winner-takes-it-all" comparison of "activation levels of behaviors", e.g., Velásquez 1997). Such implementations miss the point of *affective states as properties of the whole system that influence the behavior of the whole system at any given time*. Not surprisingly, the attribution of "affect" to such systems is usually stipulated, not argued for.

The investigations proposed in this paper are a start. Many more experiments using different kinds of affective states are needed to explore the space of possible uses of affective states and the state of possible affective states itself. Adding recurrent weights in the hidden layer, for example, would allow emotional states to influence motivational states (and vice versa) in a more direct way (e.g., the hunger drive could be suppressed by strong fear, or increasing hunger could keep the anger level up). Another direction would be to employ a more sophisticated body model (e.g., with an artificial hormone system similar to Cañamero 1997) to increase the number of controllable parameters and hence open up room for other affective states to evolve. Finally, switching to sexual procreation would facilitate more interaction among agents (e.g., competing for mates), providing yet another dimension for affective states to take control.

Acknowledgements

This research was partly funded by the Leverhulme trust and inspired by helpful discussions about affect and agent control with Aaron Sloman (who also commented on a draft version of this paper), Brian Logan and Carl Frankel.

References

Arkin, R. C. 1989. Motor schema-based mobile robot navigation. *Intern. Journal of Robotic Research* 8:92–112.

Breazeal, C. 1998. Regulating Human-Robot Interaction using 'emotions', 'drives', and facial expressions. In Proceedings of Autonomous Agents 98. Minneapolis, MO.

Cañamero, D. 1997. Modeling Motivations and Emotions as a Basis for Intelligent Behavior. In *Proceedings of the First International Symposium on Autonomous Agents*. Marina del Rey, CA: The ACM Press.

Griffiths, P. 1997. *What Emtions Really Are: The Problem of Psychological Categories*. Chicago: Chicago Univ. Press.

Meyer, J. et al. eds. 2000. *Proceedings of the Sixth International Conference on the Simulation of Adaptive Behavior*. Cambridge, MA: MIT Press.

Kauffman, S. 1995. *At home in the universe: The search for laws of complexity*. London: Penguin Books.

Maes, P. 1991. A bottom-up mechanism for behavior selection in an artificial creature. In Proc. 1st Int'l Conf. on Simulation of Adaptive Behavior, 238-246. MA: MIT Press.

Menczer, F. and Belew, R. K. 1996. From Complex Environments to Complex Behaviors. *Adaptive Behavior* 4 (3/4):317–363.

Picard, R. 1997. *Affective Computing*. MA: MIT Press.

Pfeiffer, R. 1988. Artificial Intelligence Models of Emotion. In Hamilton et al. eds. *Cognitive Perspectives on Emotion and Motivation*, 287-320. Dortrecht, Netherlands: Kluwer Academic Publishers.

Rumelhart, D., and McClelland, J. 1986. *Parallel distributed processing: Explorations in the microstructure of cognition*. Cambridge, Mass: MIT Press.

Scheutz, M. and Sloman, A. 2001. Affect and Agent Control: Experiments with Simple Affective States. In Proceedings of IAT 2001, World Scientific Publishers.

Seth, A. 2000. On the Relations between Behaviour, Mechanism, and Environment: Explorations in Artificial Evolution. Ph.D. diss., Univ. of Sussex.

Sloman, A. 1993. The mind as a control system. In Hookway, C. and Peterson, P. eds. *Philosophy and the Cognitive Sciences*, 69–110. Cambridge University Press.

Sloman, A. 2000a. Architectural requirements for human-like agents both natural and artificial (What sorts of machines can love?). In Dautenhahn, K. ed. *Human Cognition And Social Agent Technology, Advances in Consciousness Research*, 163–195. Amsterdam: John Benjamins.

Sloman, A. 2000b. Interacting trajectories in design space and niche space. In M. Schoenauer et al. eds. *PPSN VI LNSC* 1917, 3–16. Berlin: Springer.

Spier, E., and McFarland, D. 1998. Possibly optimal decision making under self-sufficiency and autonomy. *Journal of Theoretical Biology* 189:317–331.

Tyrrell, T. 1993. Computational Mechanisms for Action Selection. Ph.D. diss., Univ. of Edinburgh.

Velásquez, J. 1997. Modeling Emotions and Other Motivations in Synthetic Agents. In *Proceedings of the Fourteenth National Conference on Artificial Intelligence*. Menlo Park, CA: AAAI Press.