

Evolution of Affect and Communication

Matthias Scheutz

Human-Robot Interaction Laboratory

Cognitive Science Program

Indiana University Bloomington

Bloomington, IN 47406, USA

Introduction

Affect, or more precisely, *affective control*, is wide-spread in nature. From simple homeostatic control, to need-based control, to simple mood-based control, to basic and complex emotional control, and various other forms, affective control mechanisms of varying complexity underlie all behavior in animals. In humans, affective states are deeply intertwined with cognition and are an essential part of human mentality. And affective states often play a critical role in social behavior, from simple displays of prowess, to sexual attraction, to aggressive encounters, to social attachment, and many others. The challenge for cognitive science and its various defining disciplines including philosophy, psychology, artificial intelligence, and neuroscience is to explain what affective control is, what kinds of affective control occur in nature, how affective control can be implemented, and how it is implemented in biological organisms – we call this the “affect challenge”.

The “affect challenge” comes on the heels of much conceptual disagreement in psychology alone (but also in philosophy and artificial intelligence) about what affect concepts are. For example, the difference between *moods* and *emotions* has been explained in various non-exclusive ways: Ekman (1994) only sees them as differing in terms of time-scale with moods being longer-lasting than emotions, while for Davidson (1994) emotions bias actions while moods bias cognition; yet another explanation is offered by Frijda (1994) who distinguishes moods and emotions based on their intentionality, i.e., emotions have an object towards which they are directed, while moods are non-intentional states. To appreciate the extent of the disagreement, one does not even have to compare classes of affective states such as emotions or moods; it suffices to look at any of the classes itself, e.g., the class of emotions. As succinctly put by Delancey (2002, p. 3), “there probably is no scientifically appropriate class of things referred to by our term emotion. Such disparate phenomena—fear, guilt, shame, melancholy, and so on—are grouped under this term that it is dubious that they share anything but a family resemblance.” And, in fact, several authors have noted that there is not even agreement about what “basic emotions” are supposed to be (e.g., Ortony & Turner, 1990; Griffiths, 1997).

The “affect” challenge is, however, not limited to understanding affect concepts and the functional role of affective control processes instantiating these affect concepts in agent architectures. It also includes giving accounts of why affect is so pervasive in nature, and thus why certain forms of affective control might be better than other forms of control. In fact, we believe that understanding the dynamics of affective control processes and their utility for controlling and managing an agent’s body (e.g., against the backdrop of survival and procreation) will help in answering many open questions about the nature of individual and social behavior. For example, understanding the nature of affective control will help elucidate the different ways in which affective and cognitive processes (such as reasoning, problem-solving, and decision-making) can interact. Moreover, a detailed account of affective control processes in individuals will also help to

explain the dynamics and regulatory roles of emotion processes in social interactions (e.g., in aggressive exchanges). Most importantly, the utility and limits of affective control will allow us to determine, at least in part, possible evolutionary pressures leading to the evolution of higher-level cognition.

While we are currently a long way from being able to answer the above questions, we believe that it is possible to make headway on a smaller set of questions whose answers will contribute to making progress on the larger picture. For example, we can investigate whether the display of affective states will lead to better performance in cooperative tasks or better conflict resolution strategies in competitive tasks in multi-agent environments. We can also attempt to determine which affective control states will or are likely to evolve in cooperative and competitive multi-agent environments. And we can investigate the trade-offs between simple affective control and more complex deliberative control requiring representational mechanisms in the control architecture.

In this chapter, we focus on the utility of affective control states for ensuring an agent's survival in competitive multi-agent environments. We argue that for simple agents, simple affective control mechanisms can be defined that will result in high performance both in ordinary foraging tasks (e.g., searching for food) and in social encounters (e.g., competition for mates). As a result, we make the case that affective control via the transmission of simple signals can lead to social coordination that obviates the need for more complex forms of communication (i.e., symbolic communication based on systematic representational schemata). Methodologically, to avoid the conceptual difficulties associated with affect concepts, we will start by analyzing affect concepts in terms of *architectural capacities of agent architectures*, i.e., we explicate the architectural assumptions underlying a particular affect concept. Based on these architectural definitions, we can then define agent control architectures with mechanisms that can instantiate affect concepts, and we can implement these architectures in artificial agents in a simulated environment where agents have to perform survival tasks. Systematic experiments with these simulated agents and subsequent statistical analyzes of the results based on clearly defined performance measures then allow us to investigate the trade-offs between different architectures based on a variety of individual, social, and environmental factors (Scheutz, 2004c, 2004b).

The method of performing simulation experiments in order to investigate agent properties and be able to make claims about their likely evolution has been termed *synthetic ethology* (MacLennan, 1991) and was originally used to study the evolution of communication in a multi-agent artificial simulation environment. In a similar vein, the term “synthetic psychology” was coined (Braitenberg, 1984) to refer to the method of designing, building, and observing the behavior of artificial agents (“vehicles”) in order to study psychological principles, in general, and the difference between observable behavior and mechanisms that bring about that behavior, in particular. In our study of the evolution of affect and communication, we will use a combination of both methodologies together with a third dimension that one could call “synthetic philosophy” – the method of attempting to understand and define mental concepts, in particular affect concepts, in terms of properties of agent architectures (and environmental processes, to the extent that they are part of the control loop).

We start by introducing a very general notion of affective control, followed by a description of how simple affective control states can be implemented in an agent control architecture for simple agents performing foraging and conflict resolution tasks in a competitive multi-agent environment. We then summarize results from a large number of experiments we have performed over one decade that together provide strong evidence for the utility of simple affective control in simple organisms and thus the likely evolution of affective control systems. At the same time, we summarize results about the utility of communication showing that “representational”

communication (i.e., communication that uses systematic symbolic representations to encode meanings) is unlikely to evolve for simple organisms.

Background

Different forms of affective control have been studied over the years. Pfeifer (1988) and Sloman and Croucher (1981) are among the early investigations in artificial intelligence that laid the foundations for what is now known as “affective computing” (Picard, 1997). Researchers have started to study “affective” user interfaces, “believable” synthetic characters and life-like animated agents with emotions, affective or affect-aware instructional and virtual agents, and others (e.g., see Trappl, Petta, & Payr, 2001 for a more recent overview). While the motivations for the various research directions and their specific aims are naturally quite different (e.g., making animated characters more believable by endowing them with emotional facial expressions vs. understanding the utility of affect processes for action selection), there is a tacit assumption common to all the different approaches that affective control, in one form or another, can be useful for and might have important applications in artificial agents. We have previously compiled a list of twelve potential roles of emotions for artificial agents (Scheutz, 2004e):

1. *alarm mechanisms* (e.g., fast reflex-like reactions in critical situations that interrupt other processes)
2. *action selection* (e.g., what to do next based on the current emotional state)
3. *adaptation* (e.g., short or long-term changes in behavior due to affective states)
4. *learning* (e.g., affective evaluations used for reinforcement learning)
5. *motivation* (e.g., creating motives as part of an emotional coping mechanism)
6. *social regulation* (e.g., using emotional signals to achieve social effects)
7. *goal management* (e.g., creation of new goals or reprioritization of existing ones)
8. *strategic processing* (e.g., selection of different search strategies based on overall affective state)
9. *memory control* (e.g., affective bias on memory access and retrieval as well as decay rate of memory items)
10. *information integration* (e.g., emotional filtering of data from various information channels or blocking of such integration)
11. *attentional focus* (e.g., selection of data to be processed based on affective evaluation)
12. *self model* (e.g., affect as representations of “what a situation is like for the agent”)

This list, although not intended to be exhaustive, does point to the varied functional nature of affect mechanisms, from architectural roles to roles in social regulation. While surprisingly little work has focused on investigating roles 7 through 12 (although there are exceptions, e.g., (Gratch & Marsella, 2004)), work on interfaces, user interactions, human-robot interactions and other human-computer interaction fields has focused primarily on the sixth role, social regulation. In artificial intelligence, most work has been concerned with the first five roles, in particular, attention has been given to affective or emotional action selection, both in simulated agents (e.g. Gadanho, 2003) and robotic agents (Murphy, Lisetti, Tardif, Irish, & Gage, 2002). Similarly, quite

a bit of work has investigated the utility of evaluations that are internally generated and reflect some aspect of the internal environment (rather than the external environment) for reinforcement learning (although most reinforcement learning models in artificial intelligence do not call the evaluations “affective”). For example, Grossberg’s general *CogEM* models (e.g., Grossberg & Schmajuk, 1987) of learning cognitive, emotional, and motor properties can account for several effects in Pavlovian fear conditioning (e.g., secondary conditioning or attentional blocking), but have not been directly applied to empirical data (e.g., data from fear conditioning studies with rats). Other more targeted models of the amygdala, which performs several functions in emotion processing, assume a dual pathway model of emotional processing and were tested in auditory fear conditioning and simulated lesion studies (LeDoux, 1996). There are also attempts to include simulated bodily processes in artificial agents, e.g., simulated hormones for emotional control (Cănamero, 1997).

One difficulty with affect models is that it is often not clearly articulated why these are models of affect and what particular affective states they are models of (which sometimes leads to labels of implemented states that are familiar from human psychology such as “surprise” when, in fact, the implemented state in the model is nothing like what one would consider human surprise, e.g., Scheutz, 2002a). Rather than leaving it up to the interpretation of the reader (and risking misinterpretation), we have tried to be very specific about kinds of states we were concerned with in our work. Specifically, in Sloman, Chrisley, & Scheutz (2005) we introduced a critical high-level distinction between two very different kinds of control states in an agent architecture, those of *desire-like* and *belief-like* states, to be able to single out a general class of states that we took to be what we ordinarily intend by “affective”, namely desire-like states:

A desire-like state D of a system *S* is one whose function it is to get *S* to do something to preserve or to change the state of the world—which could include part of *S* (in a particular way dependent on *D*). Examples include preferences, pleasures, pains, evaluations, attitudes, goals, intentions, and moods.

Contrast this to belief-like states:

A belief-like state B of a system *S* is one whose function is to provide information that could, in combination with one or more different sorts of desire-like states, enable the desire-like states to fulfill their functions. Examples include beliefs (particular and general), percepts, memories, and fact-sensor states.

Affective states, being desire-like states, are thus at the heart of what causes an organism to act, either to maintain a state or to change a state. Specifically, depending on whether the organism is trying to preserve a condition, as compared to changing it, allows us to distinguish two broad classes of affective states, namely *positive* and *negative* affective states. More specifically,

A state P of a system *S* is a *positively affective state* if being in *P* or moving towards being in *P* changes the dispositions of *S* so as to cause processes which increase the likelihood of *P* persisting, or which tend to produce or enhance processes that bring *P* into existence or maintain the existence of *P*.

Again, in contrast

A state N of a system S is a *negatively affective state* if being in N or moving towards being in N changes the dispositions of S so as to cause processes which reduce the likelihood of N persisting, or which tend to resist processes that bring N into existence.

Given this simple conceptual apparatus, we can now define different affect concepts in terms of architectural mechanisms and investigate ways to implement them in agent architectures. By construing affect concepts as intrinsically architecture-based, it is possible to account for similarities and subtle differences of a great variety of affect concepts that are labeled the same in ordinary language and in large parts of the literature. A question like “Is system S capable of having *fear*?” now becomes “Is system S capable of having $fear_{16}$?”, where $fear_{16}$ is a “fear concept” specified in terms of functional capacities of an agent architecture. Such a move will not only eliminate ambiguities in language usage (“your agent model does not implement fear, it implements anxiety”), but also allow us to say when affect concepts of a particular kind are implemented (and can be instantiated) in an agent, as we are able to check whether the architecture controlling the agent supports the affect concepts specified functionally in terms of architectural requirements.

In general, for us to be able to understand the nature of affective states as they occur in biological systems, we need to (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. This questions can be widened from biological to artificial systems by adding the question (4) how affect mechanisms can be incorporated into agent architectures and implemented in artificial 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). Successful implementations of AI models, on the other hand, which employ typically simple affective states to control the behavior of simulated or real agents, provide partial answers to questions (3) and (4), but do not answer questions (1) or (2).

We believe that answers to these questions 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 that 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 trade-offs 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, 2000) 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.

We now start with a more detailed illustration of how one can define agent architectures that can instantiate affect concepts in terms of control components and then proceed to defining specific architectures for biologically plausible survival tasks that we can use to answer questions about the evolution of affect and communication.

Simple Affective Control

First and foremost, we need *control elements*, for affect as construed above essentially deals with the control of agents. Secondly, the control elements need to be connected to other components in a

way that allows them to influence the behavior and/or behavioral dispositions of the controlled agent. Finally, the *control loop* consisting of one or more control elements, the controlled components and the environment needs to be such that the functional description of the affect concepts matches the dispositional characteristics determined by the architectural layout as well as the interactions of the agent with its environment. In the following, we will focus on two rough classes of affective states, *motivational* and *emotional* states, and construe their states in general control-theoretic terms.

“Motivations” (in their most general functional description) are *desire-like states* in that they influence and bias an agent’s behavioral dispositions in such a way as to contribute to the realization of a state of the environment (including the agent) desired by the agent. They are caused by the disparity between an agent’s desire state and the state of the environment, and are themselves causes for actions that are intended to change the state of the environment so as to make it agree with the agents’ desires. Whether a state of an agent’s control system (as determined by its architecture) is a motivational state, then, depends essentially on the state’s causal connection to environmental states and its potential to influence the agent’s actions.

By “simple motivational states” we intend to refer to motivational states that have little to no cognitive involvement and are primarily linked to “basic needs” of an agent (e.g., to maintain a certain charge level of the battery in a robot). For some of these states, the familiar term “drive” is appropriate, namely if the agent is driven in a mostly reactive way to act so as to eliminate the disparity between a desired and an actual state that was the cause for the motivational state (e.g., the adjustment of a homeostatic variable). For example, for a state of an agent’s control system to qualify as a “hunger state” (i.e., an instance of a simple motivational state), roughly speaking the instantiation of this state needs to be caused by lack of energy, and needs to cause, in turn, food-seeking behavior (e.g., Lorenz & Leyhausen, 1973; McFarland, 1981). However, not all drives need to be linked to a “disparity” (e.g., play drives in dogs or cats may be triggered by boredom and thus may not be directly linked to any homeostatic imbalance), in which case they will not be (entirely) driven by motivations.

It is then possible to use outputs from gain controllers to influence motor control circuits to implement the kind of control system that will be able to instantiate such motivations. A “hunger state”, for example, could be instantiated by a *proportional controller* P (e.g., Özbay, 2000) in the following way: input to P comes from an internal sensor S_e that measures the current energy level. P compares a set point e_{des} (i.e., the desired energy level), to the actual energy level e_{act} and scales the difference by a gain factor g_e : $P = g_e \cdot (e_{des} - e_{act})$. The output then is a measure of the urgency with which the system requires energy. Hence, the intensity of the motivation is modeled by the magnitude of the output of the control circuit.

To be able to instantiate a hunger state, the controller P needs to be connected to components that control the agent’s effectors in such a way that a positive output can influence and bias the agent’s behavior towards food-seeking, where the intensity with which the agent searches for food depends on the magnitude of the output of P (reflecting the urgency with which food is needed).

Similar to motivational states, “emotions”, in the most general functional description, are *desire-like states*. They, too, influence and bias an agent’s behavioral dispositions. Yet, different from motivational states, which are linked to a disparity between a desired and an actual state, they do not have to be caused by any disparity (between an actual and a desired state). Furthermore, they themselves can be the states that the agent does or does not desire (whereas motivational states are directed towards or away from what the agent desires). A “fear state”, for example, in and of itself is an undesirable state of an agent in that it indicates (potential) danger. As such, it causes the agent to behave in such a way as to prepare for or avoid the danger. Hence, while it can also be

motivational in the sense that it may move the agent away from the cause of fear—this is the *desired state* with respect to the motivational state instantiated by a fear state—it is also emotional as it itself is not a desired state (i.e., it is a negatively valenced state). A fear state with no clearly discernible danger present, which causes an agent to be more cautious and alert, may itself not instantiate a motivational state that is connected to a particular goal such as running away from a particular threat (i.e., a desired state of the world such as “run”). Furthermore, depending on the length of duration, such a state may be better construed as an instance of a mood state in that moods lack objects at which they are directed, contrary to emotional states (although the boundary between them is not clear-cut: it is not clear when exactly a fear state, i.e., something that is caused by a perception of a dangerous situation or object, turns into an anxiety-ridden mood-like state). Emotional and motivational states are, therefore, distinct, and emotional states may or may not instantiate motivational states.¹

A simple fear state caused by the presence of dangerous objects in the environment, for example, could change the agent’s behavioral dispositions in such a way as to make it keep a certain distance from these objects for a while. Note that, while the agent is changing its behavioral dispositions insofar as it is inclined to stay away from the cause of the fear, the dispositions are changed without making the agent achieve a particular goal (i.e., desired state), e.g., to get to an environment without these objects. In fact, it may be not be possible (or if possible, a bad move) to attempt to switch from a fear state to another (more pleasant or positively valenced) state in circumstances where the source of fear can either not be pinpointed or not be avoided altogether, and general caution over an extended period of time is a beneficial (if not the only) option.

Again, we can use outputs from gain controllers to implement a control system that can instantiate an emotional state such as the above-described simple form of fear. Specifically, the fear state can be instantiated by a controller C , which integrates over time the frequency of occurrence of fear triggering conditions: input to C comes from an internal sensor S_f that is activated (under normal circumstances) by a fear triggering condition (e.g., the sensor outputs a unit impulse, Özbay, 2000). C integrates these inputs over time and outputs a signal that corresponds to the intensity of “fear”, hence to the degree with which the system should change its behavioral dispositions to be more alert, action-ready, etc. The controller could, for example, have the response characteristic given by $g(t)=e^{-t}$ to a unit impulse generated by the sensor or the perceptual system detecting a dangerous stimulus.

An example of a controller for a simple fear mechanism is given by the following differential equation:

$$\frac{\partial Act}{\partial t} = S_f \cdot G_{sensor} \cdot Act + G_{decay} \cdot Act$$

where Act is the current activation level of the control system, G_{sensor} is the gain for the sensor input and G_{decay} is the discount value for past activations. Note that the decay here is important to

¹ Note that we construe motivations to be directed towards a specific goal. In the case of an elevated “anxiety level”, as present in anxiety disorders, this kind of goal appears to be absent. Given that the state of anxiety itself is not desired, the agent could be viewed as “motivated” to change this state. If motivation is construed in this wide sense, then many emotional states may also be a motivational states (see also Ferguson, 1982). This does not seem to apply to all emotional states, however, in particular not to the kinds of states Sloman calls “tertiary emotions” or “perturbances” (Sloman, 1992), which could result from a loss of control of higher-level resource management processes in the agent’s control system.

allow the agent’s behavioral dispositions return to normal when there have not been any triggers of fear in a while.

To be able to instantiate a fear state, the above controller C needs to be connected to components that control the agent’s effectors in such a way that the positive output from C can influence and bias the agent’s behavior towards avoiding or attempting to avoid dangerous objects, where the intensity with which the agent avoids or attempts to avoid these objects depends on the magnitude of the output of C (reflecting the agent’s level of fear).

Biologically Plausible Affective Agents

The examples of motivational and emotional control mechanisms in the previous section were purposefully kept as simple as possible, for even though there is a large number of possible, much more complex affective control mechanisms that can be implemented in control circuits, we are here only interested in the simplest of control circuits that can instantiate affect concepts. For, as we will argue, these simplest of all control circuits are sufficient to allow simple agents to achieve their basic goals (of gathering food and surviving long enough to have offspring). Specifically, we now define a simple biologically inspired agent architecture that allows agents to achieve high levels of performance in foraging and conflict resolution tasks. The former task deals with the search for food, the latter with conflicts that arise when two or more agents contest a resource, given that resources are often scarce. Both tasks take place in the context of evolutionary survival tasks where agents have to collect resources in able to procreate and pass on their genes. This is because evolution discovered affective control mechanisms in the very same context of competition for survival. In fact, some neuroscientists believe that affects can be considered the major “emotional operating systems” that are defined by genetically coded neural circuits and the interactions among them (Panksepp, 1992, 1998).

We start with the basic perceptual system where a percept is viewed as generating “virtual forces” given by vector F . Force vectors point either in the direction of the percept or in the exact opposite direction depending on whether the agents perceives the object as *attractive* or *repulsive*. Moreover, the length $|F|$ determines the degree to which the agent is attracted to or repelled by the object. For example, if the perceived object is food, the agent will be generally attracted to it, but the strength of attraction might depend on how badly the agent needs food (i.e., how “hungry” the agent is). Similarly, if the object is another agent, a predator say, then the agent is repelled by it and the degree of repulsion will depend on the properties of the enemy (i.e., how strong, fast, etc., it is).

Each perceptual force vector is then scaled by the distance (typically by using $1/d^n$ with d being the distance of the agent to the object and $n>1$ depending on the type of signal drop-off, typically $n=2$ for 3d-space or $n=3$ for the plane for most signals). Scaled perceptual force vectors for each kind of object are then summed and scaled again by a time-invariant or time-variant “gain value” to account for the interest the agent has in the particular modality. More formally, the overall summed force F_{tot} is given by

$$F_{tot} = \sum_m g_{m,\varphi} \cdot \sum_{i_m=1}^{j_m} F_{i_m} / d_{i_m}^{n_m}$$

where $g_{m,\varphi}$ is the (time-invariant or time-variant) gain value of the m -th modality possibly depending on φ , F_{i_m} is the force vector to the i -th object in modality m , d_{i_m} is the distance of the agent from the i -th object in modality m , and n_m is the drop-off exponent for the m -th modality. If

g_m is a time-invariant, then the interest the agent has in modality m is the same regardless of the circumstances and ϕ is irrelevant. More often, however, $g_{m,\phi}$ will be time-variant and thus depend on other factors (e.g., time, energy levels, etc.) in which case ϕ stands for those parameters that vary over time to generate different gain values. For example, it might be that $\phi=c/l_e$ where l_e is the energy level of the agent and the gain value indirectly depends on the energy stored in the agent. In that case, an agent will have higher $g_{m,\phi}$ values when its stored energy is low and thus have higher interest in food (relative to the other modalities). The vector sum over all summed modality force vectors will then reflect this change in interest in that it will overall bias the summed force vector F_{tot} in the direction of the higher gain values, in that case the direction of food.

The total force vector is used by the agent to determine the direction in which it should go to satisfy its needs. Whether or not it will end up going in that direction, however, will depend on additional factors, e.g., on whether it has enough energy to move there, whether there is an obstacle in the way, etc. The agent's action system is the place where the final decision of what direction to move in is made. In the case of our simple agents, only available energy and physically possible motion are considered (although others are possible) to determine the speed and direction of an agent movement. Together, the perceptual and action system form the basic agent architecture for agents that can perform the foraging task. For one-resource foraging tasks, the foraging model then has two parameters: g_m and n_m for the gain and drop-off values of modality m , respectively. Each additional modality adds two more such parameters. And if there are time-variant gains, they might add more.

When resources are scarce (food, mates, territory, etc.) and multiple agents are interested in the same resource, conflicts naturally arise over those resources that need to be resolved. These contests typically involve various displays of aggression or prowess (e.g., Lorenz, 1977; Adamo & Hanlon, 1996; Hofmann & Schildberger, 2001). These expressions (e.g., facial expressions, gestures, etc.) can be construed as signals that *communicate the probability with which an animal will (continue to) fight*, where – roughly speaking – the strength of the displayed expression is directly related to the likelihood that the animal will keep fighting. Hence, for the conflict resolution task, we add another component that can determine what an agent should do if a resource is contested, i.e., when two or more agents want to eat the same food source. We also make agents display their *action tendency* and allow them to perceive the action tendency of other agents. The conflict model consists of one parameter to determine the agent's action tendency in a conflict situation: whether the agent wants to contest the resource and *fight* or whether the agent wants to move away from the resource and *flee*. We can express the agent's action tendency as the probability to fight $P(action=FIGHT)$ and define a variety of different conflict policies. For agents who ignore the display action tendencies of other agents, we can define the following policies:

- *Timid agents* never fight: $P(action=FIGHT)=0$
- *Aggressive agents* always fight: $P(action=FIGHT)=1$
- *Asocial agents* play a mixed strategy: $P(action=FIGHT)=p$ with $0 < p < 1$

Note while agents might play a fixed policy over their life-time, they could also change it based on circumstance, e.g., depending on whether they were able to win previous contests or on how low their energy level is. For example, an agent could update its policy to become “more aggressive” (i.e., increase $P(action=FIGHT)$) to be able to increase the likelihood that it will get lucky in the

future. We call such agents *adaptive* and add another parameter to the agent model that specifies how these agents change their policies over time. E.g., it might be possible for an agent to play the timid strategy until its energy level drops below a certain threshold, at which point it will start playing the aggressive strategy. Similarly, an agent might keep a tally of how many conflicts it has lost in the recent past and then increase or lower its aggression level based on how successful it was (e.g., mimicking what seems to happen in dominance relationships in social groups in nature).

Since agents will always display their action tendencies and since action tendencies are correlated with and thus predictive of the agent's actual choice and subsequent action (i.e., a high probability of fighting will often lead to "fight" decisions and outcomes), using perceived action tendencies to make decisions about one's own behavior can be beneficial. Hence, we can define several "social policies" that take both an agent's own and the perceived action tendencies into account:

- *Social agents* play a mixed strategy that depends on both their own and the perceived action tendency: $P(\text{action}=\text{FIGHT})=f(p_s, p_o)$ where $f(p_s, p_o)$ is a policy depending on the agent's own (p_s) and the other agent's perceived (p_o) action tendencies with $0 < p_s, p_o < 1$ ²
- *Rational agents* play the limit-case social strategy with $P(\text{action}=\text{fight})=1$ if $p_s > p_o$, $P(\text{action}=\text{FIGHT})=0$ otherwise

We can also imagine that agents might play different strategies not only based on their own and their opponent's (perceived) action tendencies, but also on whether they perceive the other agent to be of their own versus another agent kind. We will call those agents "discriminating agents" and allow them to use two different strategies depending on whether the contesting agent is of their own or another kind. Finally, we have so far assumed that agents will display their action tendencies truthfully, but that does not have to be the case. Hence, we will also allow for agents to cheat and lie with their displays about their true action tendencies. Such "liar agents" will display one action tendency, but act on another. E.g., an agent might display that it is maximally aggressive $P(\text{action}=\text{FIGHT})=1$, but really play a mixed strategy $P(\text{action}=\text{FIGHT})=p_s$ with $0 < p_s < 1$.

The preceding foraging and conflict models combined allow agents to forage for food and resolve conflicts that arise in multi-agent environments when multiple agents contest the same resource. The final step is to add a procreation model which allows agents to have offspring. While models of various sophistication are possible, we only consider the simplest possible model here, where an agent will automatically produce an offspring asexually when its energy level is above a given *procreation threshold*. The energy for the new agent (which is less than the procreation threshold) will be subtracted from the parenting agent's energy store and a new agent will appear close proximity to the parent agent. The child agent will inherit all parent agent's traits, with the possibility of some of the traits being mutated (based on a predefined mutation rate on the trait). This way evolutionary processes can be defined that result from genetic adaptations over time in the context of the population dynamics of surviving agents.

² For a possible function f that has the rational agent strategy as a limit case, see (Scheutz & Schermerhorn, 2004b).

Evidence for the Utility of Simple Affective Control

We are now in position to use results from simulation experiments with the affective agents defined above to make claims about the utility of affect control mechanisms and social signaling in the context of biologically plausible survival and procreation tasks that contain foraging and conflict subtasks. Specifically, we will base our claims on the success of different agent types, and consequently that of their architectures, as measured in terms of *the average number of surviving agents of a kind after a large number of agent generations* to establish that an agent kind is more likely to evolve – this is the standard evolutionary idea of *fitness*, i.e., that animals from a fitter species are more likely to procreate and pass on their genes than animals from a less fit species, which in turn might lead to the extinction of the less fit species when resources are scarce. In fact, we have been able to demonstrate that performance evaluations that measure foraging efficiency can be predictive of population dynamics and even evolutionary adaptations (Scheutz & Schermerhorn, 2005b).³

The general setup for all studies we will discuss below was a simulated unlimited 2D environment where agents have to forage for food in order to survive and procreate. Initially, specified numbers of agents and food sources are placed in the environment according to a given distribution (e.g., random, uniform, Gaussian, etc.) and then the simulation is run as a *discrete-time simulation* where, at the beginning of each simulation cycle, every agent gets to sense its environment and then decide on an action. All intended actions are then executed in parallel (with the possibility of an action failing if its enabling conditions are not given anymore). Since multiple agents are in the same environment and food is often scarce, conflicts over food can arise, hence agents have to determine whether they want to engage in a conflict over a food item or leave the scene (conflicts can also happen about other agents, but we are not pursuing this direction here). Simulations are initialized with all initial parameters fixed and then run for a certain number of steps or until some termination criterion is reached (e.g., no more agent is alive). Then different variables in the simulation environment are used for measuring agent performance (e.g., the number of surviving agents, the overall energy stored in agents, etc.). Performance measures are averaged over a set of initial conditions that are taken to be samples from a large space of initial conditions. The averages can then be used to perform various statistical analyses (ANOVAs, ANCOVAs, MANOVAs, etc.) in order to determine the dependence of performance on a set of control, bodily, social, and environmental parameters (e.g., Schermerhorn & Scheutz, 2006, 2007b). Instead of reporting the details of the experimental setup together with the specific statistical results, we will for space reasons concentrate on higher-level summaries of our findings, referring the reader to the respective publications for details. This will allow us to generate a summary of a variety of related studies that together provide strong evidence for the claims we would like to advance.

In Scheutz & Sloman (2001), we demonstrated that simple motivational agents (with “hunger-like” and “thirst-like” control mechanisms) are likely to evolve from basic agents under many environmental variations such as the distribution and influx of energy and water sources in the environment as well as the number and distribution of other agents and obstacles. These “hunger-like” and “thirst-like” states were implemented by simple feedback control circuits connected to agent-internal energy-level and water-level sensors and mutation was allowed to operate on the output of these controllers to influence the way in which the control signal was used. In all evolutionary runs, the control output evolved to be used to implement positive time-variant

³ There are, however, other measures that can be used for variants of the survival task (Scheutz & Schermerhorn, 2003).

gain values for force vectors pointing to food and water sources. Hence, the control circuit increased the agent's likelihood of moving towards food or water based on its needs, thus warranting the labels "hunger-like" and "thirst-like". Similarly, we demonstrated in Scheutz (2001) that simple "fear-like", but not "anger-like" states are likely to evolve, where the labels "fear-like" and "anger-like" were warranted because the agents evolved time-variant gain values for force vectors pointing to other agents and obstacles, thus causing them to move either away from or towards other agents and/or obstacles. We also showed that some of these connections can be learned during an agent's lifetime using simple associative learning mechanisms (Scheutz, 2000). These results were replicated and extended later in more systematic studies considering larger environmental variations (Scheutz, 2004e, 2004a).

We also investigated the trade-offs between simple reactive agents (with time-invariant gains), simple affective agents (with time-variant gains), and a third class of "deliberative" agents of varying complexity that were able to plan routes through the environment in order to acquire resources more efficiently. We showed that very simple deliberative mechanisms do not pay off in terms of overall performance, especially not if *relative performance* is considered, i.e., performance where the processing cost of using architectural mechanisms is taken into account – note that relative performance is ultimately what matters for evolutionary considerations because animals will need to spend energy for building, using, and maintaining any additional control circuits in their brains (Scheutz, Sloman, & Logan, 2000; Scheutz & Logan, 2001; Scheutz & Schermerhorn, 2002). These results were replicated in a variety of experiments investigating *foraging efficiency* in multi-agent territory exploration (MATE) tasks where a group of agents needed to collect as many resources either as quickly as possible or until all resources were collected (Scheutz & Schermerhorn, 2003, 2004a, 2004b, 2005a, 2005b; Scheutz, 2004d). Specifically, we used MATE tasks to investigate the trade-offs among architectural mechanisms, sensory range, agent group size and environmental complexity. We found that complex control systems with (optimal or close-to-optimal) planning capabilities do not pay off in environments with low structure in the distribution of food sources when relative performance is considered. In fact, not even simple "predictive mechanisms" that attempt to anticipate which food item other agents are targeting lead to better relative performance in various tasks (e.g., Schermerhorn & Scheutz, 2007a), although they can sometimes lead to better absolute performance (Scheutz & Schermerhorn, 2005a). Collectively, these studies show that simple reactive agents that perform a "greedy" search are highly effective in foraging for food. Together with the previous results about the likely evolution of simple affective control states we can conclude that affective control states will likely evolve for foraging tasks given the low architectural cost of implementing them (e.g., often a simple neuron can implement time-variant gain control). The question, then, is whether affective social control using the simple signaling model we introduced in the previous section is sufficient for coordinating social groups, as compared to more sophisticated methods of communicating information.

In a first attempt to investigate the utility of signaling internal states to other agents, we showed that taking other agents' truthfully-displayed internal fear states into account can lead to significantly better performance in multi-agent foraging tasks where conflicts can arise over resources, as compared to groups that do not indicate their fear levels (Scheutz, 2002b). Later, we designed a general game-theoretic framework for conflict resolution in simple agents and showed that there are *fair* conflict resolution strategies (for a particular notion of *fairness*) that lead to Pareto-optimal behavior (Schermerhorn & Scheutz, 2003). Moreover, we showed that there were simple ways of implementing fair strategies based on keeping track of how often an agent won or lost a conflict in the past and making one-shot behavioral decisions about who should get a resource based on this tally, effectively playing an *adaptive rational* strategy. Such adaptive

rational agents were superior to all other social and asocial agents in terms of the number of surviving agents after a certain number of generations in the conflict task (Scheutz, 2004d; Scheutz & Schermerhorn, 2004a). When agents are allowed to cheat, however, i.e., when they can wrongly indicate their action tendencies, then all truth-telling strategies will suffer (Scheutz & Schermerhorn, 2004b), which might be a reason why affective control in nature seems to be largely “hard-coded” to prevent organisms from cheating. We also analyzed the interactions between simple non-social affective control and social control through affective displays and conflict resolution strategies in order to determine the trade-offs between individual and social strategies and found that agents could make up for suboptimal strategies in the conflict task using specific gain values (behavioral propensities) in their foraging control that allowed them to avoid conflicts more frequently (Scheutz, 2006), thus providing evidence for the utility of simple (non-social) affective control (in the foraging task) in the light of conflicts, also possibly providing a way for agents to cope with cheaters in conflict tasks.

Discussion

In sum, we found strong evidence that simple affective control mechanisms implemented in terms of feedback controllers based on internal and external sensors (e.g., energy-level sensors or action tendency sensors) whose outputs are used to modulate (time-variant) gain values of percept force vectors (e.g., food sources or other agents) lead to significantly superior performance compared to time-invariant gain values, and that such control circuits are likely to evolve even in the context of competitive multi-agent environments with limited resources. Hence, the question arises whether there are circumstances in which either more complex deliberative mechanisms or more complex forms of communication are likely to evolve. We already mentioned evidence that more complex deliberative mechanisms (e.g., for planning trajectories through the environment to more efficiently collect food sources) do not have better relative performance than the considered affective mechanisms, and are thus not likely to evolve for simple agents in survival tasks (with foraging and conflict subtasks) in environments with low structure in the distribution of food sources. But this does not exclude the possibility of more complex control systems evolving in environments with more structure or in agents with more bodily limitations such as severely limited sensory ranges.

Given that more complex deliberative control systems evolved in nature, it seems clear that certain bodily, task, and/or environmental features must have provided enough evolutionary pressure for deliberative control mechanisms to evolve. For example, different from our simulations where food sources were stationary, it is likely that agents who need to deal with moving food sources (e.g., in a predator-prey scenario) will require more complex control systems for foraging and survival (e.g., to predict where prey is located/hiding, to anticipate the prey’s evasive moves in a chase, etc.). While this is clearly an important direction to pursue, it is outside the scope of this chapter. Rather, we will examine the question whether more complex forms of communication would benefit the simple kinds of agents we have defined previously, given that we know that different forms of communication have evolved in nature for different purposes, in addition to signaling action tendencies in conflict situations as discussed above (e.g., signaling danger, indicating readiness for mating, reporting locations of food). Especially since we know that humans are capable of complex symbolic forms of communication, we believe that determining the evolutionary pressures for different forms of communication to evolve is particularly interesting in light of the results about affective control systems we have reviewed so far. In fact, given that signaling action tendencies alone can already achieve a very high level of coordination and performance in simple affective agents, it is not clear whether more complex forms of

communication could significantly increase the performance of simple agents, especially if *relative performance* is considered.

For one, simple signals emitted from another agent such as “I see food here” or “I am likely going to fight” do not require much more in terms of processing on the receiving end (a simple perceptual system that can determine the signal’s direction, quality and intensity is sufficient). Contrast such “deictic signals” (Perconti, 2002) with more complex messages of the form “Agent *A* sees food *F* at location *X* at time *t*”, which require a much more complex architecture because agent names, food types, locations and times have to be explicitly decoded from the message signals (typically assuming a systematic generative signal system). Clearly, such architectures come at an additional cost (for having, maintaining, and using the additional mechanisms). Moreover, the costs of communicating can also be substantive; an agent that needs to send signals at regular intervals over a given period of time might use up a significant portion of its energy reserves, possibly without any benefit if no other agent can hear the signals (for a longer version of this argument, see Scheutz & Schermerhorn, 2008). It should be clear, then, that claims about the evolvability of communication, in particular those about the *likelihood* of communication evolving, need to be very specific about the forms of communication they target – encoding direction and distance to food sources as well as food quality requires very different (sophisticated) mechanisms from simply “annotating the environment” using pheromones like ants do.

We have investigated the utility of two major forms of communication: simple signaling, as described before, and more complex “representational messaging” where components of a communicated message require systematic representations of aspects of the environment (e.g., locations, food types, etc.). In various simulation studies, we compared simple agents with time-invariant gains with and without using signals for attracting (or repelling) other agents to (from) food sources in MATE tasks to agents that use more complex ways of communicating the location of food (“representational communication”). Overall, we found that representational communication is often not necessary to coordinate the behaviors of multiple agents in a social group (Schermerhorn & Scheutz, 2006) and that simple non-communicative predictive mechanisms can often significantly improve agent performance (Scheutz & Schermerhorn, 2005a), while adding representational communication to agents with and without this prediction mechanism does not improve the performance of the respective group. Moreover, we found that, in more structured environments where food occurs in clusters and communication does lead to better absolute performance, there is no performance difference between simple signaling and representational communication (Scheutz & Schermerhorn, 2008). And finally, even if the task complexity is increased so that multiple agents have to be at the same location in an environment at the same time, there are various simple non-communicative mechanisms that allow agents to coordinate and that lead to performance equal to, if not superior to, agents using representational communication (Schermerhorn & Scheutz, 2007a).

Conclusion

In this paper, we introduced a general notion of “affective states” that construed affect concepts as architecture-based concepts defined in terms of control mechanisms in agent architectures. We gave various examples of simple affective control states and defined several classes of simple affective agents for biologically plausible foraging and conflict resolution tasks in the context of a more general survival task. Based on the ideas of “synthetic ethology, psychology and philosophy”, we reviewed the results of various previous studies with simulated agents and, based on this evidence, arrived at the conclusion that simple affective control mechanisms are very effective at guaranteeing agents a high absolute *and*, more importantly, relative performance in the survival task in competitive multi-agent environments with little structure. Consequently, we can

claim that simple affective control states are very likely to evolve. In fact, simple affective control states turned out to be so efficient (in terms of the performance gains relative to the added cost of the control mechanisms compared to the performance of agents with time-invariant gains without additional costs) that the effectiveness of affective control is likely the cause for why evolution produced so many simple creatures that use only simple forms of signaling, if any (e.g., as measured in terms of the number of species or biomass). Comparatively, very few species were able to evolve more sophisticated control systems, including representational communication.

At present, it is not clear how and in what circumstances these more sophisticated control systems paid off relative to simpler affective control systems. One possibility is that constraints imposed on foraging tasks allowed agents to exploit them and evolve intermediate mechanisms that promised immediate performance gains. For example, we were able to show recently that the evolution of simple memories can lead to significant performance gains in foraging tasks where agents have to collect food and return it to a hive (Schermerhorn & Scheutz, 2009, 2005). Once such memories are in place, for whatever reason, agents will be able to use them in various ways. For example, they might be able to form associations between events and their affective evaluations of those events such that agents will be able to retrieve a memory item together with its affective value. This, in turn, might allow them to change gain values of force vectors in a more independent fashion compared to the fixed, determined way in which gain values are changed based on internal or external sensors. At present, this is all speculation, but it should be possible, at least in principle (if it turns out to be technologically infeasible due to an overly large number of parameters), to evaluate and ideally substantiate or refute such speculations in simulation studies with agents whose architectural parameters are systematic in the way we have done in the past. Such studies will not only contribute to a better understanding of the possible space of agent architectures and their performances, but also to a better understanding of our own mental concepts and their possible functions, including their evolutionary roles and, thus, the likelihood that instances of them would have evolved in nature.

References

- Adamo, S. A., & Hanlon, R. T. (1996). Do cuttlefish (cephalopoda) signal their intentions to conspecifics during agonistic encounters? *Animal Behavior*(52), 73–81.
- Braitenberg, V. (1984). *Vehicles: Experiments in synthetic psychology*. Cambridge, Mass: The MIT Press.
- C n mero, D. (1997). Modeling motivations and emotions as a basis for intelligent behavior. In L. Johnson (Ed.), *Proceedings of the first international symposium on autonomous agents (agents'97)* (pp. 148–155). New York, NY: ACM Press.
- Davidson, R. J. (1994). On emotion, mood, and related affective constructs. In (Ekman & Davidson, 1994) (pp. 56–58).
- DeLancey, C. (2002). *Passionate engines: What emotions reveal about mind and artificial intelligence*. Oxford University Press.
- Ekman, P. (1994). Moods, emotions, and traits. In (Ekman & Davidson, 1994) (pp. 56–58).
- Ekman, P., & Davidson, R. J. (Eds.). (1994). *The nature of emotion: Fundamental questions*. New York: Oxford University Press.
- Ferguson, E. (1982). *Motivation: An experimental approach*. Malabar, FL: Krueger Publishing Company.
- Frijda, N. H. (1994). Varieties of affect: Emotions and episodes, moods, and sentiments. In (Ekman & Davidson, 1994) (pp. 56–58).
- Gadanh , S. C. (2003). Learning behavior-selection by emotions and cognition in a multi-goal robot task. *Journal of Machine Learning Research*, 4(Jul), 385-412.
- Gratch, J., & Marsella, S. (2004). A domain-independent framework for modeling emotion. *Journal of Cognitive Systems Research*, 5(4), 269–306.
- Griffiths, P. (1997). *What emotions really are: The problem of psychological categories*. Chicago: Chicago University Press.
- Grossberg, S., & Schmajuk, N. (1987). Neural dynamics of attentionally-modulated pavlovian conditioning: Conditioned reinforcement, inhibition, and opponent processing. *Psychobiology*, 15, 195–240.

- Hofmann, H. A., & Schildberger, K. (2001). Assessment of strength and willingness to fight during aggressive encounters in crickets. *Animal Behavior*(62), 337–348.
- LeDoux, J. (1996). *The emotional brain*. New York: Simon & Schuster.
- Lorenz, K. (1977). Aggressivitt: arterhaltende eigenschaft oder pathologische erscheinung? *Aggression und Toleranz*.
- Lorenz, K., & Leyhausen, P. (1973). *Motivation and animal behavior: An ethological view*. New York: Van Nostrand Co.
- MacLennan, B. (1991). Synthetic ethology: An approach to the study of communication. In C. G. Langton, C. Taylor, J. D. Farmer, & S. Rasmussen (Eds.), *Artificial life II: Proceedings of the second workshop on artificial life* (pp. 631–658). Redwood City, Calif.: Addison-Wesley.
- McFarland, D. (1981). *The oxford companion to animal behavior*. Oxford: Oxford University Press.
- Murphy, R. R., Lisetti, C., Tardif, R., Irish, L., & Gage, A. (2002). Emotion-based control of cooperating heterogeneous mobile robots. *IEEE Transactions on Robotics and Automation*, 18(5), 744-757.
- Ortony, A., & Turner, T. (1990). What's basic about basic emotions? *Psychological Review*, 97, 315-331.
- Özbay, H. (2000). *Introduction to feedback control theory*. London: CRC Press.
- Panksepp, J. (1992). A critical role for 'affective neuroscience' in resolving what is basic about emotions. *Psychological Review*.
- Panksepp, J. (1998). *Affective neuroscience-the foundations of human and animal emotions*. Oxford: Oxford University Press.
- Perconti, P. (2002). Context-dependence in human and animal communication. *Foundations of Science*, 7, 341–362.
- Pfeifer, R. (1988). Artificial intelligence models of emotion. In V. Hamilton, G. H. Bower, & N. H. Frijda (Eds.), *Cognitive perspectives on emotion and motivation, volume 44 of series d: Behavioural and social sciences* (p. 287-320). Netherlands: Kluwer Academic Publishers.
- Picard, R. (1997). *Affective computing*. Cambridge, Mass, London, England: MIT Press.
- Schermerhorn, P., & Scheutz, M. (2003). Implicit cooperation in conflict resolution for simple agents. In *Agent 2003*. Chicago, IL: University of Chicago.
- Schermerhorn, P., & Scheutz, M. (2005, June). The effect of environmental structure on the utility of communication in hive-based swarms. In *Ieee swarm intelligence symposium 2005* (pp. 440–443). IEEE Computer Society Press.
- Schermerhorn, P., & Scheutz, M. (2006, May). Social coordination without communication in multi-agent territory exploration tasks. In *Proceedings of the fifth international joint conference on autonomous agents and multiagent systems (AAMAS-06)* (pp. 654–661). Hakodate, Japan.
- Schermerhorn, P., & Scheutz, M. (2007a). Investigating the adaptiveness of communication in multi-agent behavior coordination. *Adaptive Behavior*, 15(4), 423–445.
- Schermerhorn, P., & Scheutz, M. (2007b, April). Social, physical, and computational tradeoffs in collaborative multi-agent territory exploration tasks. In *Proceedings of the first ieee symposium on artificial life* (pp. 295–302).
- Schermerhorn, P., & Scheutz, M. (2009, March/April). The impact of communication and memory in hive-based foraging agents. In *Proceedings of the 2009 IEEE symposium on artificial life* (pp. 29–36).
- Scheutz, M. (2000). Surviving in a hostile multiagent environment: How simple affective states can aid in the competition for resources. In H. J. Hamilton (Ed.), *Advances in artificial intelligence, 13th biennial conference of the canadian society for computational studies of intelligence, ai 2000, montréal, quebec, canada, may 14-17, 2000, proceedings* (Vol. 1822, pp. 389–399). Springer.
- Scheutz, M. (2001). The evolution of simple affective states in multi-agent environments. In D. Cañamero (Ed.), *Proceedings of AAI fall symposium* (pp. 123–128). Falmouth, MA: AAI Press.
- Scheutz, M. (2002a). Agents with or without emotions? In R. Weber (Ed.), *Proceedings of the 15th international flairs conference* (pp. 89–94). AAI Press.
- Scheutz, M. (2002b). The evolution of affective states and social control. In C. K. Hemelrijk (Ed.), *Proceedings of international workshop on self-organisation and evolution of social behaviour* (pp. 358–367). Monte Verità, Switzerland.
- Scheutz, M. (2004a). An artificial life approach to the study of basic emotions. In *Proceedings of cognitive science 2004*.
- Scheutz, M. (2004b). A framework for evaluating affective control. In *Proceedings of the ace 2004 symposium at the 17th european meeting on cybernetics and systems research* (p. 645-650).
- Scheutz, M. (2004c). How to determine the utility of emotions. In *Proceedings of AAI spring symposium 2004* (p. 122-127).
- Scheutz, M. (2004d). On the utility of adaptation vs. signalling action tendencies in the competition for resources. In *Proceedings of aamas 2004* (pp. 1378–1379). ACM Press.
- Scheutz, M. (2004e). Useful roles of emotions in artificial agents: A case study from artificial life. In *Proceedings of AAI 2004* (pp. 31–40). AAI Press.

- Scheutz, M. (2006, June). Cross-level interactions between conflict resolution and survival games. In *Proceedings of artificial life x* (pp. 459–465).
- Scheutz, M., & Logan, B. (2001). Affective versus deliberative agent control. In S. Colton (Ed.), *Proceedings of the aish'01 symposium on emotion, cognition and affective computing* (pp. 1–10). York: Society for the Study of Artificial Intelligence and the Simulation of Behaviour.
- Scheutz, M., & Schermerhorn, P. (2002). Steps towards a theory of possible trajectories from reactive to deliberative control systems. In R. Standish (Ed.), *Proceedings of the 8th conference of artificial life* (pp. 283–292). MIT Press.
- Scheutz, M., & Schermerhorn, P. (2003, October). Many is more but not too many: Dimensions of cooperation of agents with and without predictive capabilities. In *Proceedings of ieeewic iat-2003* (pp. 378–384). IEEE Computer Society Press.
- Scheutz, M., & Schermerhorn, P. (2004a). The more radical, the better: Investigating the utility of aggression in the competition among different agent kinds. In *Proceedings of sab 2004* (pp. 445–454). MIT Press.
- Scheutz, M., & Schermerhorn, P. (2004b). The role of signaling action tendencies in conflict resolution. *Journal of Artificial Societies and Social Simulation*, 1(7).
- Scheutz, M., & Schermerhorn, P. (2005a). Many is more: The utility of simple reactive agents with predictive mechanisms in multiagent object collection tasks. *Web Intelligence and Agent Systems (in Press)*, 3(1), 97–116.
- Scheutz, M., & Schermerhorn, P. (2005b, June). Predicting population dynamics and evolutionary trajectories based on performance evaluations in alife simulations. In *Proceedings of gecco 2005* (pp. 35–42). ACM Press.
- Scheutz, M., & Schermerhorn, P. (2008, August). The limited utility of communication in simple organisms. In *Proceedings of artificial life xi* (pp. 521–528).
- Scheutz, M., & Sloman, A. (2001). Affect and agent control: Experiments with simple affective states. In N. Zhong, J. Liu, S. Ohsuga, & J. Bradshaw (Eds.), *Intelligent agent technology: Research and development* (pp. 200–209). New Jersey: World Scientific Publisher.
- Scheutz, M., Sloman, A., & Logan, B. (2000). Emotional states and realistic agent behaviour. In P. Geril (Ed.), *Proceedings of gameon 2000, imperial college london* (pp. 81–88). Delft: Society for Computer Simulation.
- Sloman, A. (1992). Prolegomena to a theory of communication and affect. In A. Ortony, J. Slack, & O. Stock (Eds.), *Communication from an artificial intelligence perspective: Theoretical and applied issues* (pp. 229–260). Heidelberg, Germany: Springer.
- Sloman, A. (2000). Models of models of mind, in *Proceedings Symposium on How to Design a Functioning Mind AISB'00*, Birmingham, April 2000. In (pp. 1–9).
- Sloman, A., Chrisley, R., & Scheutz, M. (2005). The architectural basis of affective states and processes. In J. Fellous & M. Arbib (Eds.), *Who needs emotions? the brain meets the machine*. New York: Oxford University Press.
- Sloman, A., & Croucher, M. (1981). Why robots will have emotions. In *Proc 7th int. joint conference on AI* (pp. 197–202). Vancouver.
- Trapp, R., Petta, P., & Payr, S. (Eds.). (2001). *Emotions in humans and artifacts*. MIT Press.