Moving Up the Food Chain:

Motivation and Emotion in Behavior-Based Robots

Ronald C. Arkin

Mobile Robot Laboratory

College of Computing

Georgia Institute of Technology

Atlanta, GA, USA 30332-0280

Phone: (404) 894-8209

Fax: (404) 385-4251

arkin@cc.gatech.edu

SHORT TITLE: MOTIVATION AND EMOTION IN BEHAVIOR-BASED ROBOTS

ABSTRACT

This article investigates the commonalities between motivations and emotions as evidenced in a broad range of animal models, including humans. In particular, a focus is placed on how these models can have utility within the context of working robotic systems. Behavior-based control serves as the primary vehicle through which emotions and motivations are integrated into robots ranging from hexapods to wheeled robots to humanoids.

Starting from relatively low-level organisms, such as the sowbug and praying mantis, and then moving upward to human interactions, a progression of motivational/emotional models and robotic experiments is reported. These capture a wide set of affective phenomena including social attachment, emotional behavior in support of interspecies interaction, multiscale temporal affect, and various motivational drives such as hunger and fear.

1. Introduction

It has been a while since I have had to wrestle with the nebulous, relatively unscientific, term "emotions". In (Arkin, 1998), I stated that "Modifying Associate U.S. Supreme Court John Paul Stevens' famous quotation, we can't define emotion, but we know it when we see it". Granted, significant advances have been recently made in understanding the neural underpinnings of emotional structure in humans (Dolan, 2002), where "emotions represent complex psychological and physiological states that, to a greater or lesser degree, index occurrences of value" (where value refers to "an organism's facility to sense whether events in its environment are more or less desirable.") Much of this recent work, however, is concerned with discovering the neurobiological underpinnings of emotions in humans and it is somewhat far removed from the more immediate needs of roboticists whose goal it is to design functioning, reliable artifacts in the real world.

While many scientists and philosophers choose to argue long and hard about the definitions of this term, classifying theories into "shallow" or "deep", (Sloman, this volume) most roboticists tend to be far more pragmatic and somewhat irreverent towards biology. We instead ask the question, what capabilities can "emotions", however defined, endow a robot with that an unemotional robot cannot possess? (Minsky, 1986) put a spin on this research hypothesis in stating, "The question is not whether intelligent machines can have any emotions, but whether machines can be intelligent without any emotions".

Unfortunately, the situation is even worse than stated above regarding the relevance of emotion to robotics. Most ethologists, scientists who study animal behavior in a natural setting, generally use the term “motivation” instead of “emotion”. As much of our group’s research historically has been derived from ethological studies, there is a strong tendency to continue to use this term in this article over the seemingly vague word emotions, even in the context of describing human behavior. Human behavior, at least from my perspective as a roboticist, can be characterized to a great extent through ethological studies, e.g., (Blurton Jones, 1972) that abstract away from neural models of brain function in favor of observational studies. It is also unclear, at least to me, as to where within the range of animal species the ability to possess emotions begins. Does a {paramecium, sowbug, canine, human} express emotion? These questions seem better left to others than myself, as I am unconvinced that their pursuit will lead to more intelligent robots, which some consider a new species in their own right (Menzel and D’Alusio, 2000).

Motivations, however, tend to be more general than emotions, especially when concerned with human performance (See Rolls, Kelley, this volume). They often involve the articulation of goals that result in the performance of goal-achieving behavior. Thus when pressed to define the distinction between emotions and motivations I state the following working definition (caveat - this is a roboticist speaking): emotions constitute a subset of motivations that provide support for an agent’s survival in a complex world. They are not related to the formulation of abstract goals that are produced as a result of deliberation. Motivations and emotions affect behavioral performance, but motivation can additionally lead to the formulation of concrete goal-achieving behavior, at least in humans, whereas emotions are concerned with modulating existing behaviors in support of current activity. In this regard, motivations might additionally invoke specific behaviors to accomplish more deliberative tasks or plans (e.g., strategies for obtaining food).

It is my view that motivations (and emotions) affect the underlying control of a cybernetic system by altering the underlying behavioral parameters of the agent, whether it is biological or artifactual, (i.e., a robot). Certain internal states, which are used to represent various motivation/emotional qualities, are maintained by processes that reflect the agent's time course through the environment as well as its perception of the immediate situation. Using this definition, it then becomes our goal, as roboticists, to design systems that can maintain this internal motivational state and use it to produce behavior in ways that are consistent with intelligent performance in the real world.

Motivations/emotions provide two potential crucial roles for robotics:

1. *Survivability*: Emotions serve as one of the mechanisms to complete autonomy and that helps natural systems cope with the world. (Darwin, 1972) postulated that emotions serve to increase the survivability of a system. Often a critical situation does not allow time for deliberation, and emotions modulate the behavioral response of the agent directly.
2. *Interaction:* Many robots that are created to function in close proximity to people need to be able to relate to them in predictable and natural ways. This is primarily a limitation of the human, whom we do not have the luxury of reprogramming. In order to make robots interact effectively and efficiently with people it is useful for them to react in ways that humans are familiar and comfortable with (See Norman, this volume).

This article will present a range of research results that attempts to address the issues above while spanning the phylogenetic complexity of various animal models: i.e., moving up the food chain. We first look at the lowly sowbug as a basis for incorporating motivation behavior, then move up to predatory insects, specifically the praying mantis. Moving into the realm of humans we then investigate intra-species behavior, the mother-child relationship, and then inter-species interaction in the relationship of a robotic dog with its owner. Finally we summarize a relatively complex model of motivations that includes multiple timescales formulated in terms of traits, attitudes, moods, and emotions. Hopefully the journey through these various biological entities and their robotic counterparts will help demonstrate a basis for commonality of emotion and motivation across all species, while simultaneously encouraging others to loosen their definitional belt a bit regarding just what emotions are.

2. Tolman's Schematic Sowbug and its Robotic Counterpart

Our first study looks at a psychological model of the behavior of a sowbug. Tolman introduced his concept of a schematic sowbug, which was a product of his earlier work on purposive behaviorism developed in the early 1920's. (Innis, 1999) states:

> *Initially, in Tolman's purposive behaviorism, behavior implied a performance, the achievement of an altered relationship between the organism and its environment; behavior was functional and pragmatic; behavior involved motivation and cognition; behavior revealed purpose.*

Motivation was incorporated into the tight connection between stimulus and response that the prevailing behaviorist view largely ignored While Tolman also developed the notion of the cognitive map, this was not used in his Sowbug model. Instead motivation was used to create additional inputs to his overall controller, something that more traditional behaviorists tended to ignore. These relations were expressed (Tolman, 1951) in the determination of a behavioral response (B) as a function receiving inputs from environmental stimuli (S), physiological drive (P) (or motivation), heredity (H), previous training (T), and age (A).

(Tolman, 1939) initially proposed the concept of the schematic sowbug as a thought experiment to express the concepts of purposive behaviorism. It was infeasible at the time to think of physically constructing a robotic implementation of his Sowbug, even if he had cared to. Space prevents a detailed description of his model; and the reader is referred to (Tolman 1939, Endo and Arkin, 2001) for the details. The Sowbug was, in Tolman's view, a simple creature capable of moving autonomously through an obstacle strewn area in search of food to maintain its existence, where its performance is affected by drives, stimuli, and the other factors mentioned earlier. Described at a high-level, Tolman's schematic sowbug consisted of:

- A receptor organ: a set of multiple photo-sensors that perceive light (or any given stimuli) in the environment and mounted on the front-end surface of the sowbug, forming an arc.
- Movement was based on the following components for determining the sowbug's behavioral response:
    - Orientation distribution that indicates the output values of the photo-sensors serving as environmental stimuli.
    - Orientation/Progression tensions which correspond to the motivational demand. The term tension here refers to the readiness of the Sowbug to pursue a stimulus, the readiness in this case being derived from hunger (i.e., the hungrier it is (motivation), the greater is its readiness (tension). The orientation tension refers to its readiness to turn, while the progression tension indicates the Sowbug's readiness to move forward.
    - Orientation vector which when generated rotates the sowbug in the appropriate direction.

- Progression distribution that reflects the strength (or certainty) of a specific stimulus.
- Progression vectors that represent the velocities of the left- and right-hand side motors of the sowbug, similar to what (Braitenberg, 1984), in his book *Vehicles*, described nearly 40 years later.

The orientation/progression tensions are the key underlying mechanisms by which motivation is introduced into the model. These tensions modulate the sowbug's response to given environmental objects (e.g., food) by representing a variable for motivational drives (e.g., hunger). The result is the *orientation need*, which directly alters the strength of the agent's motor response to a given stimulus.

What is remarkable is that Tolman's schematic sowbug was the first prototype in history that actually described a behavior-based robotics architecture, to the best of our knowledge. It was a half-century before (Brooks, 1986) developed the subsumption architecture. Past training and internal motivational state also affect the sowbug's behavior. Endo (Endo and Arkin, 2001) created a partial robotic incarnation of Tolman's schematic Sowbug. His software, called eBug[1] (Emulated Sowbug), supports both simulations (Fig. 1 left) and robot experiments (Fig. 1 right) that reflect many of the details of Tolman's model.

Tolman's purposive behaviorism spoke to many of the same issues that modern behavior-based robotics architectures address (Arkin, 98): how to produce intelligent behavior from multiple concurrent and parallel sensorimotor (behavioral) pathways, how to coordinate their outputs meaningfully, how to introduce the notion of goal-oriented behavior, how to include motivation

---

[1] `EBug is available at http://www.cc.gatech.\\edu/ai/robot-lab/research/ebug/`

and emotion, and how to permit stages of developmental growth to influence behavior. Tolman's approach has yet to be fully exploited for motivational and emotional control in real robots, and indeed may never be as more modern theories of affect now exist, but the work is certainly of historical significance to both roboticists and those in psychology who have the vision to create computational models that can serve as the basis for robotic intelligence.

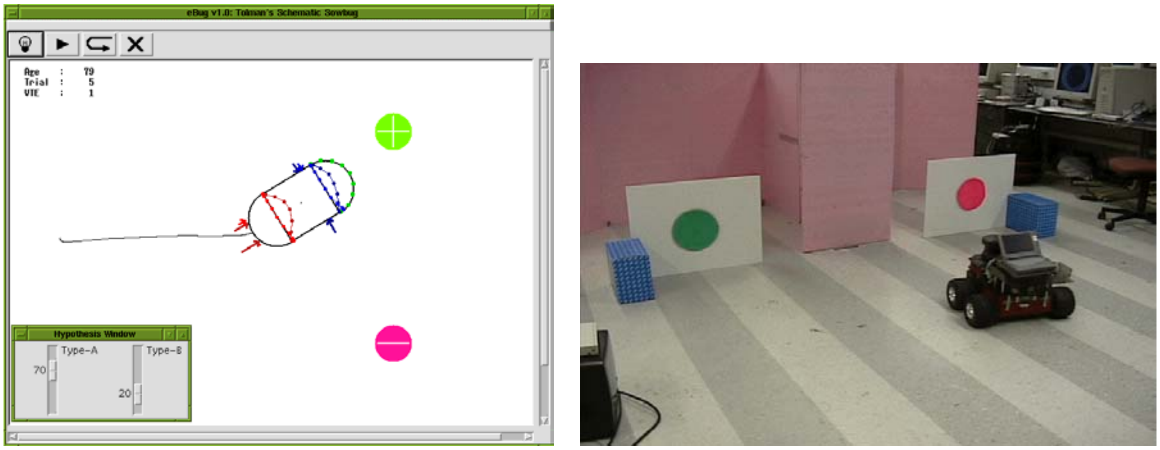

Figure 1: **(Left) eBug Simulation of Schematic Sowbug. Initially the two colored objects appear to both be potential food sources. The robot learns over time that the light colored object is an attractive (food) stimulus, while the darker object is an aversive one, and changes its behavior as a result. (Right) eBug controlling an actual robot to react to the same colored stimuli. The robot is attracted to the object on the right.**

3. Motivational behavior in Mantids and Robots

Moving on from the lowly sowbug to animals that prey on insects (i.e., the praying mantis), we strive to understand how the basic drives (motivations to ethologists) affect their behavior in an attempt to create similar models for robotic systems. One behavioral model based upon schema theory (Arbib, 1992) and earlier work applying schema theory to frog behavior (Arbib 1987) has been used to represent the insect's participation with its world. This involves the extension of our robotic schema-theoretic approach (Arkin, 1989) to incorporate internal motivational processes in addition to external perception. Fortunately, schema theory is quite amenable to this strategy for the mantis, which we have demonstrated both in simulations and in actual robotic experiments (Arkin et al., 2000). One early example of integrating motivation into navigation was forwarded by (Arbib and Lieblich, 1977) that made explicit use of drives and motivations integrated into spatial representations, and although not implemented, explained a variety of data of rats running mazes. Others (e.g., Steels, 1994; McFarland, 1993) have also explored similar issues experimentally. Our overall approach (Arkin, 1990) is also related to ecological and cognitive psychology as formulated by (Gibson, 1977) and (Neisser, 1976) respectively.

As we have seen earlier, the efficacy of visual stimuli to release a response (i.e., type of behavior, intensity, and frequency) is determined by a range of factors: the stimulus situation (e.g., form, size, velocity, and the spatio-temporal relationship between the stimulus and the animal); the current state of internal variables of the organism, especially those related to motivational changes (e.g., season of the year, food deprivation, time interval between feeding the animal and the time of experimentation) and; previous experience with the stimulus (e.g., learning, conditioning, habituation).

In a joint project between researchers at Georgia Tech and the Instituto Technológico Autónomo de México (ITAM) in Mexico City, a behavioral model was created (Arkin, Cervantes-Perez, and Weitzenfeld, 1998) that captures the salient aspects of the mantid that incorporates 4 different visuomotor behaviors (a subset of the animal’s complete behavioral repertoire):

- **Prey Acquisition**: This behavior first produces orienting, followed by approach (if necessary), then grasping by the mantis when the target is within reach.
- **Predator Avoidance**: At the most abstract level, this produces flight of the insect, but when considered in more detail there are several forms of avoidance behavior. A large flying stimulus can yield either a ducking behavior or a fight-type response referred to as deimatic behavior where the insect stands up and opens its wings and forearms to appear larger than it is.
- **Mating**: This is an attractive behavior generated by a female stimulus during the mating season that produces an orienting response in the male followed by approach, then actual mating.
- **Chantlitaxia**: [Etymology: from the Nahuatl word “Chantli” which means shelter or refuge, and “taxia” from the Latin for attraction (Cervantes-Perez, 2003)]. This involves an agent's search for a proper habitat for survival and growth. The praying mantis climbs to higher regions (e.g., vegetation) when older, actively searching for a suitable place to hunt.

This model incorporates motivational variables that affect the selection of these motivated behaviors. For predator avoidance, fear is the primary motivator; for prey acquisition, hunger serves a similar purpose; while for mating, the sex-drive dominates. These variables are modeled quite simply in this instance, but they may be extended to incorporate factors such as diurnal, seasonal, and climatic cycles and age-related factors as discussed in some of the other models that follow in this article. The behavioral controller, derived from (Cervantes-Perez et al., 1993) and depicted in Figure 2, was implemented on a small hexapod robot.

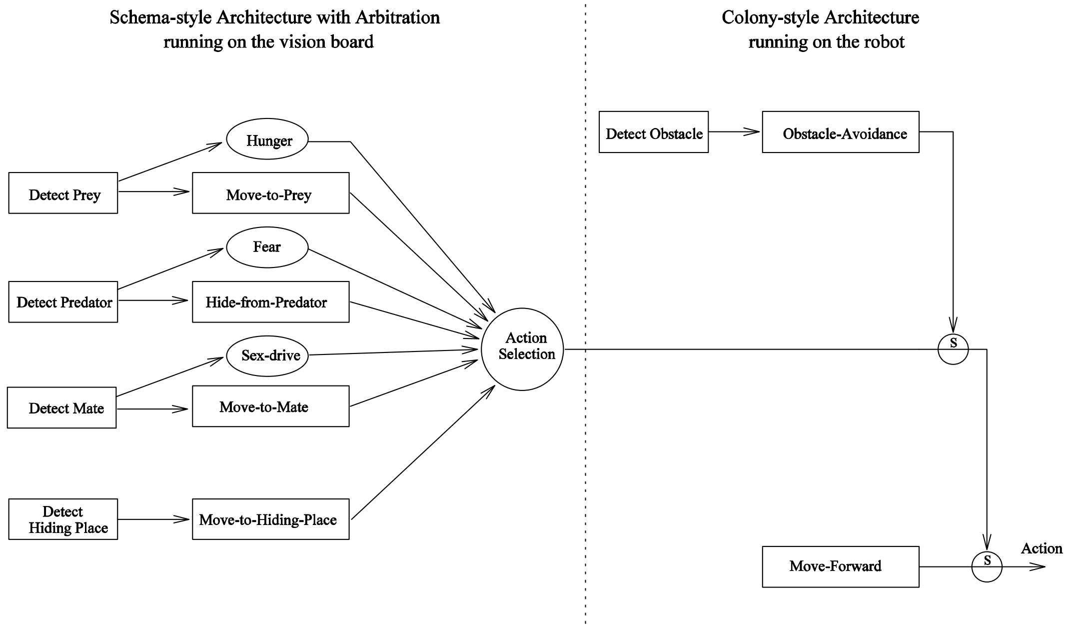

Figure 2: Implemented Behavioral Model. A schema-style architecture (left) involves fusing the outputs of the behaviors together cooperatively, while a colony-style architecture (right) involves competitive priority-based arbitration for action selection (Arkin, 1998).

In this implementation, rather than responding to movement, as is the case for the mantis, the robot responds instead to colors. Green objects represent predators, purple objects represent mates, orange objects that are at least twice as tall as they are wide represent hiding-places, and all other orange objects represent prey. The robot maintains three motivational variables that represent the robot's hunger, fear, and sex-drive. Initially, the value of each of these variables is set to zero. Arbitrarily, the hunger and sex-drive levels increase linearly with time, with the hunger increasing at twice the rate of the sex-drive. When the robot has contacted a prey or mate, the robot is considered to have eaten or mated with the object, and the relevant motivational variable resets to zero. Contact is determined by the position of the prey or mate color blob in the image captured by the camera mounted on the front of the robot. In this case, the object is considered to be contacted when the bottom of the object blob is in the lower five percent of the image. The fear level remains zero until a predator becomes visible. At that time, the fear variable is set to a predetermined high value. When the predator is no longer visible, the fear level resets to zero.

Grey Walter's turtle (Walter, 1953) had an alternative perspective for being 'motivated' avoiding the use of state variables and encapsulating it within the behaviors themselves: e.g., Avoid light unless you are hungry for a recharge. This type of motivation affects the action selection of the behaviors based on external immediate stimuli, whereas a more explicit representation format is encoded in the mantis model.

In the robotic implementation of the mantis model, motivational values directly influence the underlying behavioral control parameters, altering them in a manner consistent with the agent's needs. For example, when the robot has a high hunger level, and food appears as a stimulus, the behavior associated with moving towards food is strong. On the other hand, if the robot does not have a high hunger motivational value, it will ignore the food stimulus. Similar behavioral reactions occur for the fear motivational value in the presence of predators, and the sex drive variable when a mate is present. The behavior that dominates the overall performance of the hexapod is determined to a great extent by the internal motivational variables and is no longer solely driven by the visual stimuli present in the robot's current field of view. Figure 3 illustrates one of many examples of the robot's behavior using this model (Arkin et al., 2000). It is relatively straightforward to incorporate more complex motivational modeling if it were available, although this work was reserved for more complex species, as discussed in sections 4-6 of this article. Examples of more complex relationships might include an increase of the hunger variable after sex or perhaps a concomitant increase in fear sensitivity during sex. These relationships can in principle be determined neurologically, if not behaviorally. The relationships may change in time and could possibly be said to *define* the 'emotional' state of the mantis, if emotions were to be attributed to insects.

What this work illustrates is the ability to readily integrate a range of motivational variables, some of which might be construed as emotions (reluctantly, due to concerns over the inherent looseness of this term as mentioned in the introduction of this article) into a behavior-based architecture. The net effect is that the behaviors are no longer solely driven by external perceptions but now also by internal state. These states can be affected by elapsed time, perceived events, or other factors. Drawing on biologically inspired models of robotic control provides an easy mechanism by which motivational state can be introduced into robots that at least imitate lower-life forms in some respects. We now move forward and upward to look at how humans interact with each other and other species as a basis for capturing additional nuances of emotional behavior in robotic systems.

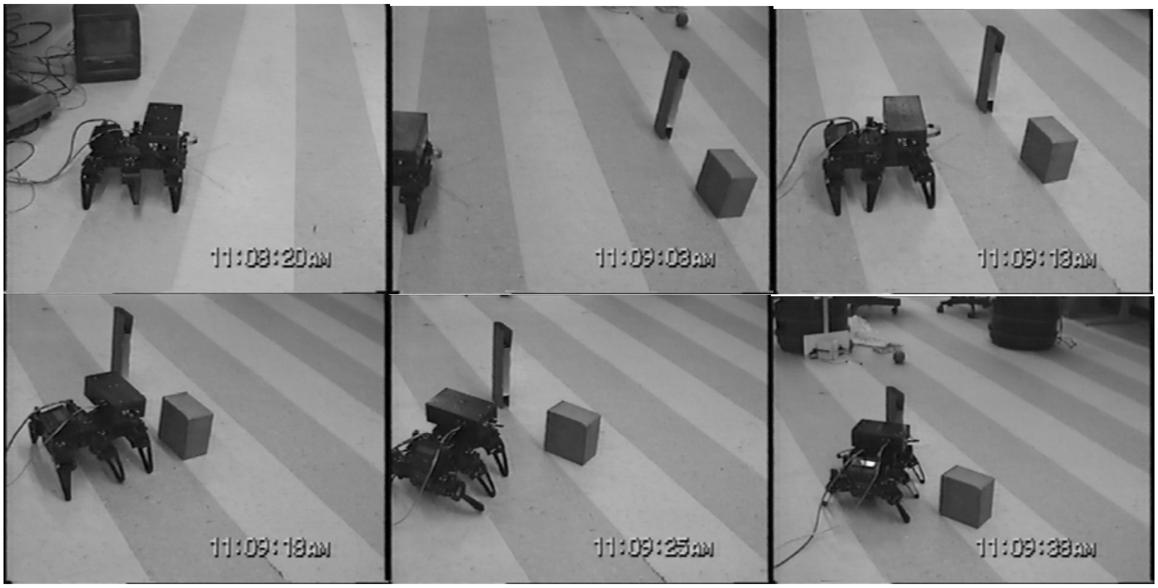

Figure 3: Starting at the upper left and moving along the top, the robot is initially still due to a high fear level as a predator object is in view off to the right. Once that is removed, the robot moves towards a prey

object satisfying its need for hunger. Once its appetite has been satisfied it now ignores the food object and moves toward the tall object that represents a mate.

4. Attachment Theory as a Basis for Robot Exploration

We now investigate the relationship of parent and child as a basis for the child's emotional state and its potential for use within robotics. In particular we look at Bowlby's work (Bowlby, 1969) that focuses on emotional attachment, which is reflected in what we refer to as "comfort level". In humans, both external (exogeneous) environmental conditions and internal (endogeneous) state determine this level of comfort, reflecting the perceived degree of safety in the current environment and the degree of normal functioning of our internal system.

The input features of comfort consist of these two components, at least for infants (Dunn, 1977). Endogenous factors include hunger, body temperature, pain, and violent or sudden stimulation received by any of the infant's sensors. One of the most significant exogeneous factors is environmental familiarity. Hebb's Discrepancy Theory (Hebb, 1946) states that fear and discomfort are evoked by events that are very different from previous experiences. (Dunn, 1977) elaborates that whether or not the past experience with the current situation was pleasant, neutral, or unpleasant is significant. An infant brings to the evaluation of any situation a predisposition threshold on whether to react with pleasure or fear (Stroufe, 1974).

(Bowlby, 1969) created a theory of attachment in which he pointed out that infants associate certain individuals (caregivers) with security and comfort. They use these people as sources of comfort. In their early years, children want to maintain close proximity to their caregiver, and the degree to which they want to maintain this proximity depends on the circumstances. "The behavioral hallmark of attachment is seeking to gain and to maintain a certain degree of proximity to the object of attachment, which ranges from close physical contact under some circumstances to interaction or communication across some distance under other circumstances." (Ainsworth and Bell, 1970). The mother-child relationship is the primary exemplar of this interaction where the child prefers to be close to the mother in unfamiliar situations, especially when young. Every attachment object has an attachment bond between itself and the child, where the force of the attachment is situationally dependent and is directed towards reducing the distance of separation between the child and the attachment object (Bowlby, 1969).

Likhachev (Likhachev and Arkin, 2000) extrapolated these ideas to working robotic systems. Taking some liberty with formal attachment theory, we now view that an infant (robot) maximizes its exogenous and maximum endogenous comfort components by being physically collocated with its mother (object of attachment). The attractive force is a function of the attachment bond that corresponds to the object, the individual's overall comfort level, and the separation distance.

The intent here is not to create a robotic model of a human child, but rather to produce useful behavior in autonomous robotic systems. While robots are not children, there are nonetheless advantages to maintaining an attachment bond with certain individuals or objects, (e.g. caregivers, owners, military base, fuel supply, or familiar end-users), as they typically satisfy the robot's endogenous needs (e.g., energy) while also providing a high-level of familiarity and predictability in the environment. Each attachment object has an associated attachment bond. The degree to which the robot bonds to a particular object depends on how its needs are met by that object and the level of comfort it can achieve.

3.1. Computational Model of Attachment

The result of attachment behavior is a response directed toward increasing or maintaining proximity with the attachment object (Colin, 1996). In a schema-based model, this results in an attractive vector directed toward the object of attachment of varying magnitude dependent upon the separation distance. This vector magnitude (A) represents the intensity of the attachment, which is functionally:

$$A = f(C, \alpha, d), \tag{1}$$

where C is the overall comfort level of a robot; α is the attachment bonding quality between the robot and the particular attachment object in question, and d is the distance between the robot and the attachment object. Specifically, A is defined as the product of the normal attachment maximum level N, quality of attachment α, and the amplification of the comfort component in the function by a proximity factor D:

$$A = N * \alpha * D * \varphi(C), \tag{2}$$

The normal attachment maximum level N defines the maximum magnitude of the attachment intensity when the object of attachment is a normal "mother", so to speak. The other factors in the function, with the exception of α, are normalized. The attachment bonding quality (α) should be dependent on the quality of care that the attachment object provides for the robot, but is set arbitrarily in advance for the results reported below. Setting α to 1 corresponds to a "normal mother" attachment object. Setting α greater than 1 corresponds to "over-caring mother", whereas decreasing α below 1 corresponds to "under-caring mother". Setting α to 0 corresponds to "no-care mother", resulting in the complete absence of attachment behavior in a robot.

The relationship between A and C expressed in the comfort component φ(C) is drawn from the following two sources. (Feeney and Noller, 1996) describe comfort-seeking intensity in adults as a function of anxiety and fear that is linear for secure adults, where adults typically form attachments with parents, siblings, friends, and partners. It is similarly treated linearly for the robot experiments that appear below. (Colin, 1996) identifies a low-level attachment behavior activation where the behavior has almost no effect but only monitors the proximity, and a strong activation where the behavior's output overrides virtually all other behaviors in the system when the distance of separation becomes significant. Mathematically this relationship can be described as follows:

$$\varphi(C)=\begin{cases} A_l & \text{if } C>C_l \\ \dfrac{A_h-A_l}{C_h-C_l}*C-\dfrac{A_h-A_l}{C_h-C_l}*C_h+A_h & \text{if } C_h<C<C_l, \\ A_h & \text{if } C<C_h \end{cases} \tag{3}$$

where C_l and C_h define low and high comfort activation levels, respectively; and A_l and A_h are the corresponding intensity levels for the low and high activation levels.

The proximity factor D is a function of the distance d from the robot to the attachment object. As defined, when the robot is very near the attachment object the proximity factor is set to 0, in effect zeroing the attachment force since the robot is already sufficiently close to its object of attachment. This results in a *Safe Zone* that forms a secure area where the robot receives maximum comfort. When the robot moves outside this safe zone, the proximity factor grows, increasing the overall attachment force, until reaching a maximum at some distance. This area between the safe zone and the distance where the maximum proximity factor occurs is called the *Comfort Zone*, (Fig. 4) which constitutes the normal working region for the robot. Outside of this comfort zone the attachment force is quite large and generally forces the robot to move into and stay within its comfort zone.

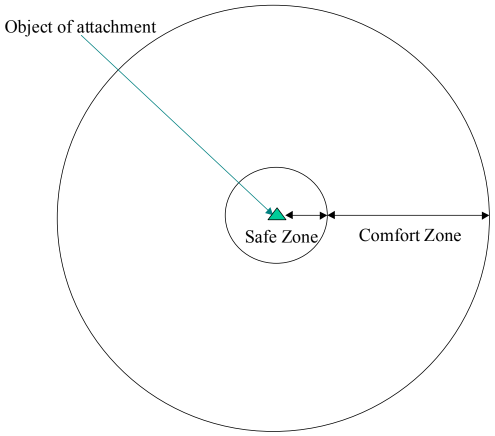

Figure 4: **The safe and comfort zones of the robot around the object of attachment.**

3.2. Robotic Experiments

Figure 5 depicts the effects of various settings of comfort levels on a simulated robot's performance during an exploration task in the presence of an attachment object. As the robot becomes less comfortable it remains closer to its object of attachment. A complete statistical analysis of these and other results are presented in (Likhachev and Arkin, 2000). Similar results were obtained during actual robotic experiments (Figure 6).

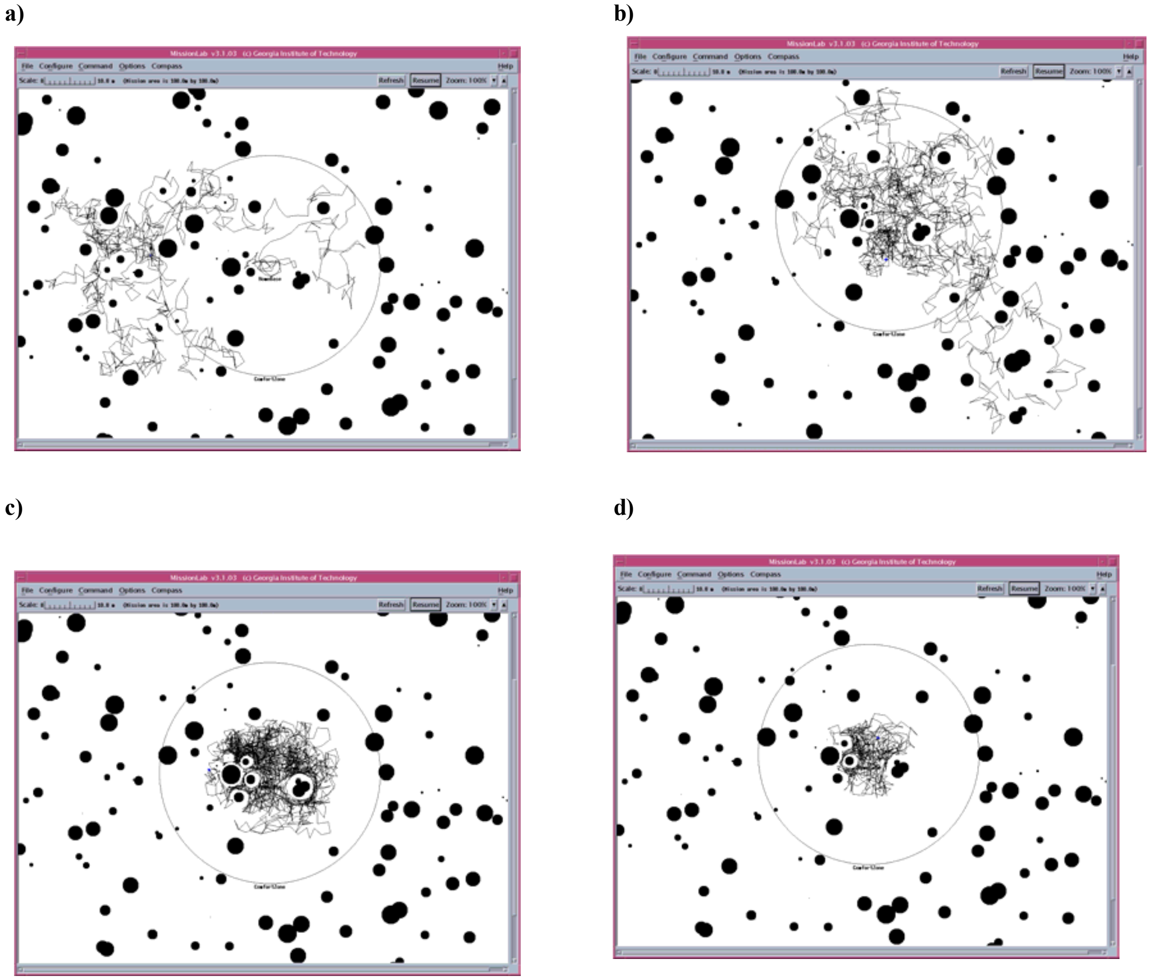

Figure 5. 3 minute runs of exploration behavior with the attachment object of attachment located at the center of the circle that defines the comfort zone. (a) No attachment behavior; (b) attachment behavior with comfort level set 1.0 (maximum comfort); c) comfort level set to 0.0 (neutral comfort); d) comfort level set to -1.0 (maximum discomfort)

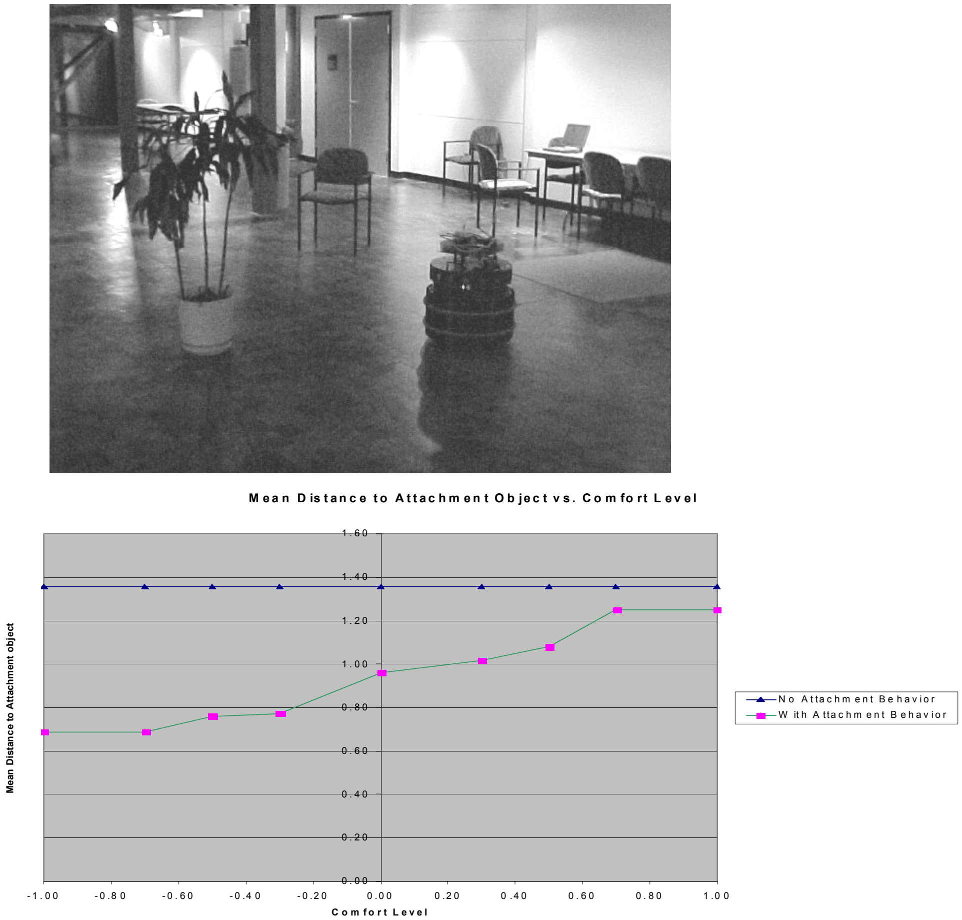

Figure 6: (Top) Nomad Robot conducting 5-minute explorations. The object of attachment is the tree. (Bottom) Results showing how average distance from attachment object increases as robot's comfort level increases.

The notion of emotional comfort as a basis for modulating behavior can have significant impact in controlling a robot's performance as it moves through the world. This is not only of value in ensuring that the robot does not stray from a given task or area with which it is familiar but can also provide a basis for establishing inter-species bonding in entertainment robotics, where creating a pet robot that can effectively relate to a human is of great importance. The next section of this article focuses on various aspects of human-robot interaction in this new application domain for robots.

5. Canine Ethology in Support of Human-Robot Bonding.

One of the principal goals of entertainment robotics is to provide the illusion of life in a robot to a human. A strategy we have chosen to follow, in joint work with Sony Corporation (Arkin et al., 2001, Arkin et al., 2003), is to develop a computational model of behavior based on ethology. In this work, we start to engage the concepts of motivational behavior in animals, specifically dogs; with that of emotionality experienced in humans. One of the goals is to produce appropriate emotional responses in people through observation and interaction with a robotic artifact. This requires the generation of natural behaviors as well as maintaining motivational/emotional states within the robot. Studies of the manifestation of emotions in humans and their similar occurrence as motivational behavior in animals can provide support for effective interactivity between a robot and a human (Breazeal, 2002; Dautenhamn, 1999; Fujita, 2001, Breazeal, this volume). By incorporating aspects of emotional and motivational behavior into a robotic architecture we, and others, (e.g., Breazeal and Scasselati, 1999), contend that a greater ability to relate to the end-user is provided.

Figure 7: Various AIBO Robots.

The primary robotic system used for this work is Sony's AIBO, a highly successful commercial product (Figure 7). A broad range of behaviors is available organized into multiple subsystems (Arkin et al., 2001). Their selection is related to the motivational state of the robot, maintained in what is referred to as the I/E (instinct/emotion) model. Ekman's model (Ekman, 1994) has been influential in this work, and consists of 6 basic emotional states: happiness, anger, sadness, fear, surprise, and disgust (cf. Rolls' dimensions (Rolls, this volume) and Ekman's dimension as illustrated by Kismet (Breazeal, this volume)). Takanishi's approach (Takanishi, 1999) is also used to reduce the overall internal state space into 3-dimensions: pleasantness, arousal, and confidence. The 6 basic emotional states are located within this 3-dimensional space. By establishing predefined levels of internal variables, such as hunger and thirst, and determining how the current state of the robot relates to those thresholds, pleasantness can be assessed. If these variables remain within the regulated range, the pleasantness is high. Arousal is controlled by both circadian rhythm and unexpected stimuli, while confidence is determined by the confidence (certainty) of recognized external stimuli.

The resulting emotional values affect the action-selection process for behavior eligibility for execution. Drawing on aspects of both McFarland's motivational space (McFarland, 1974) and Blumberg's action-selection mechanisms (Blumberg, 1994), particular behaviors are scheduled for execution on the robot that are consistent with the current set of environmental stimuli and the internal state of the robot itself. In the action selection module, a behavior is selected based on inputs derived from external stimuli (releasing mechanisms) and the robot's current motivational state variables. A state-space diagram represents the mapping from these inputs onto the appropriate behavior to be activated. The details of this approach can be found in (Arkin et al., 2003).

Further extension of this research by Fujita has resulted in the EGO architecture (Fujita et al., 2001a, 2001b) leading to potential applications in humanoid robots (Arkin et al 2003b). The generated motion patterns can be affected by the emotions themselves. (Fujita, 2001b) specifically addresses the symbol-grounding problem (Harnad, 1990) in this architecture, allowing the robot to learn to associate behaviors with specific symbols through the use of an "emotionally grounded symbol", where the physically grounded symbol is associated with the change of internal variable state that occurs when the robot applies a behavior in response to the object. For example, when the robot hears the symbol's name spoken, it knows which behavior(s) are associated with that symbol and can produce a change in the robot's internal motivations. Thus, in a sense, the robot knows the meaning of the symbol in the way in which it affects both its internal state and what behaviors are the correct ones to use in the associated object's presence. The symbols are grounded not only perceptually, by associating the correct perceptual stimuli with the spoken symbol, but also behaviorally, by producing the appropriate behavioral response in the presence of the stimuli that acts in a manner to produce a change in internal variables that are consistent with the IE model. This use of symbols for emotional modeling diverges somewhat from strict ethology, especially when compared to the more faithful canine behavioral modeling employed (Arkin et al, 2003), but keep in mind that the intent is to create robotic artifacts that successfully entertain and engender human-robot emotional bonding.

More recent work has moved into expanding this architecture into humanoid behavior for the Sony SDR-4X robot (Fig. 8), which is also capable of emotional expression. Extensions to the EGO architecture include the introduction of a deliberative layer capable of planning (Arkin et al., 2003). Proprietary issues prevent a more thorough discussion at this time.

AIBO and SDR are neither dogs nor humans. They are entertainment robots intended to provide interesting and engaging experiences with people. The use of emotions in these systems is created with that particular purpose in mind. An underlying belief is that if a robot is capable of expressing itself not only through speech but also emotionally, they are more likely to be accepted by consumers.

In all of the robots discussed thus far, internal state refers to the maintenance of a set of variables that reflect the emotional/motivation state of the machine. This appears to be consistent with Dolan's definition of emotion that appears in the Introduction, with the exception that there is relatively little complexity involved in our implementation. These variables are updated continuously by arriving sensory information, and in some cases are updated by circadian rhythms and other factors. This set of states acts on the behavioral regime of the robot to modulate and/or select behaviors that best reflect the current set of emotional conditions. Considerably more complex models are possible, one of which is discussed in the next section.

Figure 8: Sony SDR-4X manifesting a happy greeting consistent with its emotional state

6. A New Model, Summary and Conclusion

We conclude the paper with the presentation of a new model under development in our laboratory at Georgia Tech, and then summarize the overarching themes of this article.

6.1 Traits, Attitudes, Moods, and Emotions (TAME)

We are not currently aware of any single computational model that captures the interaction between a wide range of *time-varying* affect-related phenomena, such as personality traits, attitudes, moods, and emotions. In humans, each of these components performs a distinct adaptive function. It is our research hypothesis that providing autonomous robots with similar easily recognizable affective cues may facilitate robot-human in complex environments.

Moshkina (Moshkina and Arkin, 2003) has proposed a new affect model called TAME, which incorporates and unites Traits, Attitudes, Moods and Emotions as separate components with well-characterized interfaces in order to produce multi-scale temporal affective behavior for use in a behavior-based robotic system. TAME draws from a number of related theories of personality, mood, emotion and attitudes, but it is intended to serve primarily as a basis for producing intelligent robotic behavior and not as a cognitive model of affect and personality.

The Personality and Affect Module is currently being integrated into the Autonomous Robot Architecture (AuRA) (Arkin and Balch, 1997) as embodied in the *MissionLab* Mission Specification System[2] (Mackenzie, Cameron, and Arkin, 1997). The personality and affect module modifies the underlying behavioral parameters, which directly affect currently active behaviors, similar to earlier work from our laboratory in homeostatic control (Arkin, 1992) and the work described for the praying mantis in Section 3 of this article. Moshkina's conceptual view of the TAME model is presented in figure 9.

Psychologists have factored affective responses into at least these four components: Traits, Attributes, Moods, and Emotions. By attempting to apply temporal attributes to this differentiated set of affective response factors, we feel this can add clarity to the generation of affective behavior in robots through the generation of new computational models enable the composition of these time-varying patterns. This will hopefully give rise to mechanisms by which a robot's responses can be more appropriately attuned to a human user's needs, in both the short- and long-term. It is not intended to validate any particular theories of human affective processing, but rather to assist in creating better human-robot interaction.

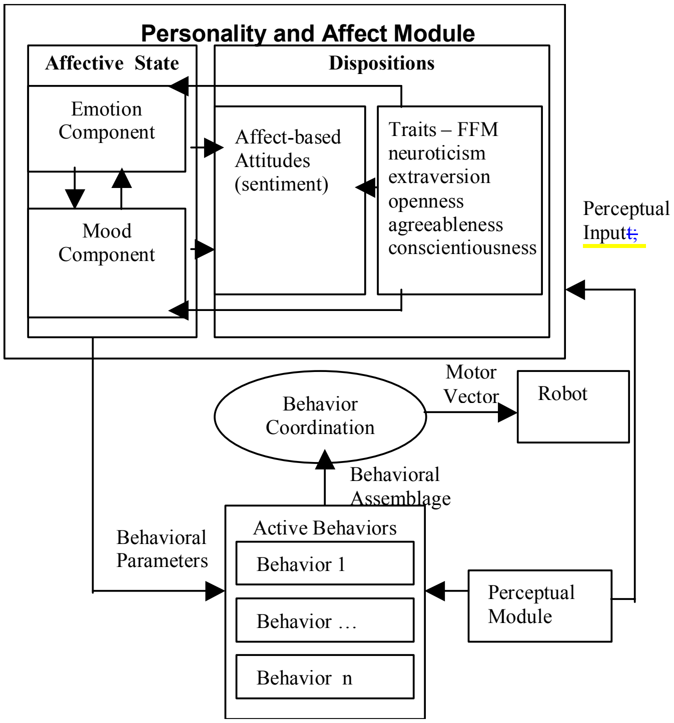

Figure 9: Integrated Model of Personality and Affect (TAME)

---

[2] MissionLab is freely available at www.cc.gatech.edu/ai/robot-lab/research/MissionLab.html

The four major components operate in different time and activation scales. Emotions are high activation and short-term, while moods are low-activation and relatively prolonged. Traits and attitudes determine the underlying disposition of the robot and are relatively time-invariant. The basis for each of these four components is discussed briefly below.

Traits serve as an adaptation mechanism to specialized tasks and environments, whereas emotions mobilize the organism to provide a fast response to significant environmental stimuli. The Five-Factor Model (FFM) of Personality developed by (McCrae and Costa, 1996) serves as the basis for the trait components. Trait dimensions include *Openness (O), Agreeableness (A), Conscientiousness (C), Extroversion (E), and Neuroticism (N).* Traits influence a wide range of behavior, and are not limited to emotionally charged situations.

Emotion, in the TAME context, is an organized reaction to an event that is relevant to the needs, goals, or survival of the organism (Watson, 2000). It is short in duration, noncyclical, and is characterized by a high activation state and significant energy and bodily resources expenditure. A typical set of emotions that we subscribe to includes joy, interest, surprise, fear, anger, sadness and disgust (Watson, 2000) and are continuously dynamically generated as emotion-eliciting stimuli are detected.

Moods bias behavior according to favorable/unfavorable environmental conditions, and are defined by the two independent categories of positive and negative affect (Revelle, 1995). They constitute a continuous affective state that represents low activation state and is less intense and thus expends less energy and bodily resources than emotion. Moods are mainly stimulus-independent, and exhibit cyclical (circadian) variation according to time of day, day of the week, and season.

An attitude is a “learned predisposition to respond in a consistently favorable or unfavorable manner with respect to a given object” (Breckler and Wiggins, 1989). Attitudes guide behavior towards desirable goals and away from aversive objects, as well as facilitate decision-making process by reducing the decision space complexity. They are relatively time-invariant, object/situation specific, influenced by affect, and result in a certain behavior towards the object.

To test the TAME model, a partial integration of the personality and affect module into the *MissionLab* system was undertaken, which is a supplemented version of the Autonomous Robot Architecture (Arkin and Balch; 1997). The interested reader is referred to (Moshkina and Arkin, 2003) for additional details. Research is now underway in the administration of formal usability studies to determine whether this form of affect can play a significant role in improving a user’s experience with a robot.

6.2 Summary and Conclusion

In the end, what can we learn from this journey through a broad range of motivations/emotions than span multiple species? I propose the following:

- Emotions, at least to a roboticist, consist of a subset of motivations that can be used to dynamically modulate ongoing behavioral control in a manner consistent with survival of the robotic agent (Arkin and Vachtsevanos, 1990). The nuances surrounding which species possess ‘emotions’ versus ‘motivations’ and the terminological differences between these terms are best left to non-roboticists in my opinion, as it is unclear if the resolution to these semantic differences will have any impact whatsoever on our ability to build more responsive machines. Our community, however, sorely needs more and better computational models and processes of affect that effectively capture these components within a behavioral setting.

- Human-robot interaction can be significantly enhanced by the introduction of emotional models that serve as much for the benefit of the human as well as the robot.
- Motivational/Emotional models can be employed that span many, many different organisms and that can match the requirements of an equally diverse robotic population ranging from vacuum cleaners, to military systems, to entertainment robots, and others. All of these systems need to survive within their ecological niche and must respond to a broad range of threats towards their extinction or obsolescence. The principle of biological economy would argue that emotions/motivation exist in biology to serve a useful purpose and it is our belief that robots can only benefit by having a similar capability at their disposal.
- The diversity of emotional models is something to celebrate and not lament, as they all can potentially provide fodder for robotic system designers. As I have often said, I would use phlogiston as a model if it provided the basis for creating better and more intelligent robots, even if it does not explain natural phenomena accurately.

Finally, there is much more work to be done. This branch of robotics has only recently been enabled due to major computational and hardware advances that have only existed within the past few decades. As such it is an exciting time to be studying these problems in the context of artificial entities.

ACKNOWLEDGMENTS

The author is deeply indebted to the researchers whose work is reported within this article including Yoichiro Endo, Khaled Ali, Francisco Cervantes-Perez, Alfredo Weitzenfeld, Maxim Likhachev, Masahiro Fujita, Tsubouchi Takagi, Rika Hasegawa, and Lilia Moshkina. Research related to this article has been supported by the National Science Foundation, DARPA, and the Georgia Tech GVU Center. The author also thanks Dr. Doi, the director of Digital Creatures Laboratory, Sony, for his continuous support for our research activity.

REFERENCES

Ainsworth, M.D.S. and Bell, S.M., 1970, Attachment, Exploration and Separation: Illustrated by the Behaviour One-Year-Olds in a Strange Situation, Child Development, 41:49-67.

Arbib, M.A., 1992, Schema Theory, in Encyclopedia of Artificial Intelligence, 2nd Edition, (S. Shapiro, Ed.) 2:1427-1443, New York:Wiley.

Arbib, M.A. and Lieblich, I., 1977, Motivational Learning of Spatial Behavior, in Systems Neuroscience, (Ed. J. Metzler), London: Academic Press, pp. 221-240.

Arkin, R.C., 1989, Motor Schema-Based Mobile Robot Navigation, International Journal of Robotics Research, 8(4):92-112.

Arkin, R.C., 1990, The Impact of Cybernetics on the Design of a Mobile Robot System: A Case Study, IEEE Transactions on Systems, Man, and Cybernetics, 20(6):1245-1257.

Arkin, R.C., 1992, Homeostatic Control for a Mobile Robot: Dynamic Replanning in Hazardous Environments, Journal of Robotic Systems, 9(2):197-214.

Arkin, R.C., 1998, Behavior-based Robotics, Cambridge:MIT Press.

Arkin, R.C., Ali, K., Weitzenfeld, A., and Cervantes-Perez, F., 2000, Behavioral Models of the Praying Mantis as a Basis for Robotic Behavior, Journal of Robotics and Autonomous System, 32 (1):39-60.

Arkin, R.C. and Balch, T., 1997, AuRA: Principles and Practice in Review, Journal of Experimental and Theoretical Artificial Intelligence, 9(2-3):175-189.

Arkin, R.C., Cervantes-Perez, F., and Weitzenfeld, A., 1998, Ecological Robotics: A Schema-Theoretic Approach, in Intelligent Robots: Sensing, Modelling and Planning, (R.C. Bolles, H. Bunke, and H. Noltemeier, Eds.), Singapore:World Scientific, pp.377-393.

Arkin, R., Fujita, M., Takagi, T., and Hasegawa, R., 2003, An Ethological and Emotional Basis for Human-Robot Interaction, Robotics and Autonomous Systems, 42(3-4).

Arkin, R.C., Fujita, M., Takagi, T., and Hasegawa, R., 2001, Ethological Modeling and Architecture for an Entertainment Robot, 2001 IEEE International Conference on Robotics and Automation, Seoul, Korea, pp.453-458.

Arkin, R.C. and Vachtsevanos, G., 1990, Techniques for Robot Survivability, Proc. 3rd International Symposium on Robotics and Manufacturing, Vancouver, BC, pp.383-388.

Blumberg, B., 1994, Action-Selection in Hamsterdam: Lessons from Ethology, From Animals to Animats 3, (D. Cliff et al., Eds.), MIT Press, pp.108-117.

Blurton Jones, N., 1972, Ethological Studies of Child Behavior, London:Cambridge Press.

Bowlby, J., 1969, Attachment and Loss, I: Attachment, London:Hogarth Press.

Braitenberg, V., 1984, Vehicles: Experiments in Synthetic Psychology, Cambridge:MIT Press.

Breazeal, C., 2002, Designing Sociable Robots, Cambridge:MIT Press.

Breazeal, C., and Scassellati, B., 1999, How to Build Robots that Make Friends and Influence People, Proc. IROS 1999, pp.858-863.

Breckler, S.J. and Wiggins, E.C., 1989, On Defining Attitude and Attitude Theory: Once More with Feeling, in Attitude Structure and Function, (A.R. Pratkanis, S.J. Breckler, and A.G. Greenwald Eds.), Mahwah, N.J.:Lawrence Erlbaum Associates, pp. 407-429.

Brooks, R., 1986, A Robust Layered Control System for a Mobile Robot, IEEE Journal of Robotics and Automation, RA-2(1):14-23.

Cervantes-Perez, F., Personal Communication, 2003.

Cervantes-Perez, F., Franco, A., Velazquez, S., Lara, N., 1993, A Schema Theoretic Approach to Study the "Chantitlaxia" Behavior in the Praying Mantis, Proc. of the First Workshop on Neural Architectures and Distributed AI: From Schema Assemblages to Neural Networks, USC.

Colin, V.L., 1996, Human Attachment, New York:McGraw-Hill.

Darwin, C., 1965, The Expression of the Emotions in Man and Animals, Chicago:University of Chicago Press.

Dautenhahn, K. and Billard A., 1999, Bringing up Robots or Psychology of Socially Intelligent Robots: From Theory to Implementation, Proc. 3rd International Conference on Autonomous Agents, Seattle, WA, pp.366-367.

Dolan, R.J., 2002, Emotion, Cognition, and Behavior, Science, 298(8 Nov. 2002):1191-1194.

Dunn, J., 1977, Distress and Comfort, Cambridge:Harvard University Press.

Ekman, P., and Davidson, R. J., 1994, The Nature of Emotion, Oxford:Oxford University Press.

Endo, Y. and Arkin, R.C., 2001, Implementing Tolman's Schematic Sowbug: Behavior-based Robotics in the 1930s, 2001 IEEE International Conference on Robotics and Automation, Seoul, Korea, pp.487-484.

Feeney, J. and Noller, P., 1996, Adult Attachment, London:SAGE Publications.

Fujita M., et al., 2001a, Physically and Emotionally Grounded Symbol Acquisition for Autonomous Robots, Proc. AAAI Fall Symposium: Emotional and Intelligent II, pp.43-46.

Fujita M., et al., 2001b, An Autonomous Robot that Eats Information via Interaction with Human and Environment, Proc. IEEE ROMAN-01.

Gibson, J.J., 1977, The Theory of Affordances, in Perceiving, Acting, and Knowing, (R. Shaw and J. Bransford, Eds.), Mahwah, N.J.:Erlbaum.

Harnard, S., 1990, The Symbol Grounding Problem, Physica D, 42:335-346.

Hebb, D.O., 1946, On the Nature of Fear, Psychological Review, 53:259-276.

Innis, N.K., 1999, Edward C. Tolman's Purposive Behaviorism, in Handbook of Behaviorism, Academic Press, pp.97-117.

Likhachev, M. and Arkin, R.C., 2000, Robotic Comfort Zones, Proc. SPIE Sensor Fusion and Decentralized Control in Robotic Systems III, pp.27-41.

MacKenzie, D., Arkin, R.C., and Cameron, R., 1997, Multiagent Mission Specification and Execution, Autonomous Robots, 4(1):29-52.

McFarland, D. and Bosser, T., 1993, Intelligent Behavior in Animals and Robots, Cambridge:MIT Press.

McFarland, D. (Ed.), 1974, Motivational Control Systems Analysis, London:Academic Press.

McCrae, R.R. and Costa, P.T., 1996, Toward a New Generation of Personality Theories: Theoretical contexts for the five-factor model, in Five-Factor Model of Personality, New York:The Guilford Press, pp.51-87

Menzel. P. and D'Alusio, F., 2000, Robo Sapiens, Cambridge:MIT Press.

Minsky, M., 1986, The Society of the Mind, New York:Simon and Schuster.

Moshkina, L., and Arkin, R.C., 2003, On TAMEing Robots, Proc. IEEE International Conference on Systems, Man and Cybernetics, Oct. 2003.

Neisser, U., 1976, Cognition and Reality: Principles and Implications of Cognitive Psychology, New York:Freeman.

Revelle, W., 1995, Personality Processes, Annual Review of Psychology, 46:295-328

Steels, L., 1994, A Case Study in the behavior-oriented design of Autonomous Agents, From Animals to Animats 3, (D. Cliff et al., Eds.), pp.445-452.

Stroufe, L.A., Waters, E., and Matas, L., 1974, Contextual Determinants of Infant Affective Response, The Origins of Fear, (M. Lewis and L. A. Rosenblum, Eds.), New York: Wiley.

Takanishi, A., 1999, An Anthropomorphic Robot Head having Autonomous Facial Expression Function for Natural Communication with Human, 9th International Symposium of Robotics Research (ISRR99), pp.197-304.

Tolman, E.C., 1939, Prediction of Vicarious Trial and Error by Means of the Schematic Sowbug, Psychological Review, (H.S. Langfeld, Ed.), 46:318-336.

Tolman, E.C., 1951, Behavior and Psychological Man, Berkeley: University of California Press.

Walter, G., 1953, The Living Brain, New York:Norton.

Watson, D., 2000, Mood and Temperament, New York:The Guilford Press.