Robot Emotion: A Functional Perspective

Cynthia Breazeal and Rodney Brooks

1 Introduction

Why should there be any serious research at all on the possibility of endowing robots with emotions? Surely this is the antithesis of engineering practice which is concerned with making functional devices rather than ones which invoke emotions in people–the latter is the realm of art or at best design.

Over the last one hundred years the average home in the Western World has seen the adoption of new technologies that at first seemed esoteric and unnecessarily luxurious. These include electricity, refrigeration, running hot water, telephone service, and most recently, wideband internet connections. Today the first few home robots are starting to be sold in conventional retail stores. Imagine the world fifty years from now when robots are common in everybody's home. What will they look like and how will people interact with them?

Electricity, refrigeration, and running hot water are utilities that are simply present in our homes. The first robots that have appeared in homes are largely also a utility, but have a presence that triggers in some people responses that are normally only triggered by living creatures. People do not name their electrical outlets and talk to them, but they do name their robots and, today at least, carry on largely one-sided social interactions with them. Will it make sense, in the future, to capitalize on the tendency of people for social interactions in order to make machines easier to use?

Today's home cleaning robots are not aware of people as people. They are not even aware of the difference between a moving obstacle and a static obstacle. So today they treat people as they would any other obstacle; something to be skirted around while cleaning right up against them.

Imagine a future home cleaning robot, fifty years from now, and the capabilities it might plausibly have. It should be aware of people as people as they have very different behaviors than do static objects such as chairs or walls. When a person happens to be walking down the hallway where a robot is cleaning it might determine that there is a person walking towards it. A reasonable behavior would be to then get out of the way. By estimating the person's mood and level of attention to their surroundings, the robot can determine just how quickly it should get out of the way. If a person points to the corner of a room and says "clean more over here" the robot should understand the focus of attention of the person as that room corner, rather than the end of their finger. If the person utters an angry "no!" directed towards the robot it should re-evaluate its recent actions and adapt its future behavior in similar circumstances.

These are all examples of the robot reading cues, some emotional, from a person. But a robot might also be easier to interact with if it provides social cues to people.

As the robot notices a person approaching down a corridor it might, like a dog would, make brief eye contact with the person (here the person has to estimate its gaze direction) and then give a bodily gesture that it is accommodating the person and the person's intent. As it is cleaning that dirty corner it might express increased levels of both frustration and determination as it works on a particularly difficult food spill on the carpet so that the person who directed its attention to that corner can rest assured that it has understood the importance of their command and that it will do what it takes to fulfill it. Upon hearing that angry "no!" the robot may express its chagrin in an emotional display so that the person understands intuitively that their command has been heard and will be reflected upon.

A robot with these sorts of capabilities would seem luxuriously out of place in or current homes. But it might be that such capabilities will come to be expected, as they will make robots more natural and simpler to interact with.

In order to get to these future levels of social functionality for a robot we need to investigate how to endow our robots with emotions and how to enable them to read social and emotional cues from people.

2 Functional Roles of Emotions

All intelligent creatures that we know of have emotions. Humans, in particular, are the most expressive, emotionally complex, and socially sophisticated of all (Darwin, 1872).

To function and survive in a complex and unpredictable environment, animals (including humans) were faced with applying their limited resources (e.g., muscles, limbs, perceptual systems, mental abilities, etc.) to realize multiple goals in an intelligent and flexible manner (Gould, 1982). Those species considered to be the most intelligent tend to exist in complex and dynamic social groups where members have to communicate, cooperate, or compete with others.

Two conceptually distinct and complementary information processing systems, cognition and emotion, evolved under such social and environmental pressures to promote the health and optimal functioning of the creature (Damasio, 1994; Izard & Ackerman, 2000). As argued by Norman, Ortony and Russell in this volume, the cognitive system is responsible for interpreting and making sense of the world, whereas the emotion system is responsible for evaluating and judging events to assess their overall value with respect to the creature (e.g., positive or negative, desirable or undesirable, etc.).

Emotion theorists agree that the cognitive and emotion systems are interrelated. One view privileges the cognitive system where the cognitive processes of appraisal and attribution recruit emotions. Others see emotion and cognition as being reciprocally interrelated, recognizing that each emotion often recruits and organizes cognitive processes and behavioral tendencies in a specific manner to the adaptive advantage of the creature (Izard, 1993). For instance, according to Izard, a unique function of sadness is its ability to slow the cognitive and motor systems. For example, Termine & Izard (1988) found that mothers' expression of sorrow through facial and vocal expression during face-to-face interactions with their 9-month-old infants significantly decreased their babies' exploratory play. In adults, the slowing of cognitive processes may enable a more careful and deliberate scrutiny of self and circumstances, allowing the individual to gain a new perspective to help improve performance in the future (Tomkins, 1963).

Numerous scientific studies continue to reveal the reciprocally interrelated roles that cognition and emotion play in intelligent decision making, planning, learning, attention, communication, social interaction, memory, and more (Isen, 2000). Emotion plays an important role in signaling salience, to guide attention toward what is important and away from distractions, thereby helping to effectively prioritize concerns (Picard, 1997). Isen has studied the numerous beneficial effects that mild positive affect has on a variety of decision making processes for medical diagnosis tasks—e.g. facilitating memory retrieval (Isen et al., 1978); promoting creativity and flexibility in problem solving (Estrada et al., 1994); and improving efficiency, organization and thoroughness in decision making (Isen et al., 1991). As argued in Isen (1999), negative affect allows us to think in a highly focused way when under negative, high-stress situations. Conversely positive affect allows us to think more creatively and to make broader associations when in a relaxed positive state.

Furthermore, scientists are finding that whereas too much emotion can hinder intelligent thought and behavior, too little emotion is even more problematic. The importance of emotion in intelligent decision making is markedly demonstrated by Damasio's studies of patients with neurological damage that impairs their emotional systems (Damasio, 1994). Although these patients perform normally on standardized cognitive tasks, they are severely limited in their ability to make rational and intelligent decisions in their daily lives. For instance, they may lose a lot of money in an investment. Whereas healthy people would become more cautious and stop investing in it, these emotionally impaired patients do not. They cannot seem to learn the link between bad feelings and dangerous choices, so they keep making the same bad choices again and again. The same pattern is repeated in their relationships and social interactions resulting in loss of jobs, friends, and more (Damasio, 1994; Picard, 1997). By looking at highly functioning autistics, we can see the crucial role that emotion plays in normal relations with others. They seem to understand the emotions of others like a computer—memorizing and following rules to guide their behavior, but lacking an intuitive understanding of others. They are socially handicapped, not able to understand or interpret the social cues of others to respond in a socially appropriate manner (Baron-Cohen, 1995).

2.1 Emotion Theory Applied to Robots

This chapter presents a pragmatic view of the role emotion-inspired mechanisms and capabilities could play in the design of autonomous robots—especially as it is applied to human-robot interaction (HRI). Given our discussion above, many examples could be given to illustrate the variety of roles that emotion-inspired mechanisms and abilities could serve a robot that must make decisions in complex and uncertain circumstances, either working alone or with other robots. Our interest, however, concerns how emotion-inspired mechanisms can improve the way robots function in the human environment, and how such mechanisms can improve robots ability to work effectively in partnership with people. In general, these two design issues (robust behavior in the real world, and effective interaction with humans) are extremely important given that many real-world autonomous robot applications require robots to function as members of human-robot teams.

We illustrate these advantages with a design case study of the cognitive and emotion-inspired systems of our robot, Kismet. We have used the design of natural intelligences as a guide, where Kismet's cognitive system enables it figure out what to do, and its emotion system helps it to do so *more flexibly* in complex and uncertain environments (i.e., the human environment), as well as to help Kismet behave and interact with people in a socially acceptable and natural manner.

This endeavor does not imply that these emotion-based or cognition-based mechanisms and capabilities must be in some way identical to those in natural systems. In particular, the question of whether or not robots could have and feel human emotions is irrelevant to our purposes. Hence, when we speak of robot emotions, we do so in a functional sense. We are not claiming that they are indistinguishable from their biological correlates in animals and humans. Nonetheless, we argue that they are not "fake" because they serve a pragmatic purpose for the robot that mirrors their natural analogs in living creatures.

Furthermore, the insights these emotion-based and affect-based mechanisms provide robot designers should not be glossed over as merely building "happy" or entertaining robots. To do so is to miss an extremely important point: as with living creatures, these emotion-inspired mechanisms modulate the cognitive system of the robot to make it *function better* in a complex, unpredictable environment—to allow the robot to make *better* decisions, to learn *more* effectively, to interact *more* appropriately with others—than it could with its cognitive system alone.

3 Emotion-Inspired Abilities in Human-Robot Interaction

There are a diverse and growing number of applications for robots that interact with people. These applications include surgery robots, scientific explorers, robots for search and rescue, surveillance and telepresence robots, museum docents, robot toys, entertainment robots, robotic prosthetics, robot nursemaids, educational robots, and more.

The demands that the robot's architecture must address depends on a number of issues. For instance, is the robot completely autonomous, teleoperated by a human, or somewhere in between? Is the robot's environment controlled and predictable, or is it complex and unpredictable, even potentially hostile? Is the robot designed to perform a specific task, or must it satisfy multiple and potentially competing goals? Is the robot expected to function in complete isolation or in cooperation with others? What is the nature of their interaction? Does the human use the robot to mediate his or her own actions, using the robot as a tool or a prosthesis? Does the human cooperate with the robot as a teammate? Does the robot provide some form of companionship, such as a pet? Is it expected to enter into long-term relationship with a particular person, such as a nursemaid?

In Breazeal (2003*d*), we classify these applications into four different paradigms of interaction (see below). Each is distinguished from the others based on the mental model a human has of the robot when interacting with it. Furthermore, for each there are a wide assortment of advantages that giving robots skills and mechanisms associated with emotion could play. These include:

- Intelligent behavior in a complex, unpredictable environment.
- The ability to sense and recognize affect and emotion.

- The ability to express affect and internal state in familiar human terms.
- The ability to respond to humans with social adeptness.

Robot as a tool. In this first paradigm, the human views the robot as a device that is used to perform a task. The amount of robot autonomy varies (and hence the cognitive load placed on the human operator) from complete teleoperation to a highly self-sufficient system that need only be supervised at the task level. Consider a specialist who uses a robot to perform tasks autonomously in complex and often hazardous environments. This might be a scientist in interaction with a robot to explore planetary surfaces, the ocean depths, etc. Alternatively, it could be a fireman working with a search and rescue robot to survey a disaster site. In both of these cases, the communication between the robot and the human is very limited (e.g., large delays in transmissions or limited bandwidth). As a result, the robot must be self sufficient enough to perform a number of tasks in difficult environments where the human supervises the robot at a task level.

Much like an animal, the robot must apply its limited resources to address multiple concerns (performing tasks, self preservation, etc.) while faced with complex, unpredictable, and often dangerous situations. For instance, balancing emotion-inspired mechanisms associated with interest and fear could produce a focused yet safe searching behavior for a routine surveillance robot. For this application, one could take inspiration from the classic example of Lorenz (1950) regarding the exploratory behavior of a raven when investigating an object on the ground starting from a perch high up in a tree. For the robot just as for the raven, interest encourages exploration and sustains focus on the target, while recurring low levels of fear motivates it to retreat to safe distances, thereby keeping its exploration within safe bounds. Thus, an analogous exploratory pattern for a surveillance robot would consist of several iterative passes toward the target: on each pass move closer to investigate the object in question and return to a launching point that is successively closer to the target.

Robot as cyborg extension. In the second paradigm, the robot is physically merged with the human to the extent that the person accepts it as an integral part of his/her body. For example, the person would view the removal of his/her robotic leg as an amputation that leaves him/her only partially whole. Consider a robot that has an intimate physical connection with one's body, such as an exoskeleton for a soldier or a prosthetic of an amputee. Emotions play a critical role in connecting the mind with the body and vice versa. Note that the performance of one's physical body changes depending on whether one is relaxed or exhilarated. Although a robotic leg would not "have emotions" itself, it is important that it adapt its characteristics to match those of the rest of the body in accordance with the emotional state of the human to avoid imbalance. If the robotic extension were able to sense and recognize the person's emotional state (perhaps via physiological changes in his/her body) it could adapt its operating characteristics appropriately. Under calmer times the robotic extension could go into energy conservation mode since power demands are lower. However, in high stress situations, the robot could change its operation parameters to significantly augment the person's strength or speed.

Robot as avatar. In this third paradigm, a person projects him/herself through the robot to remotely communicate with others (the next best thing to being there). The robot provides a sense of physical presence to the person communicating through it, and a sense of social presence to those interacting with it. Consider robot-mediated communication with others at a distance. Technology mediated communication today is rather impoverished compared to face-to-face conversation, limiting our diverse communication channels to a select few. The advantage and appeal of a robot avatar is the ability to have a more fully embodied experience for the user , and a greater physical and social presence to others (including touch, eye contact, physical proximity, movement and gesture within a shared space, etc.). The cognitive load and physical coordination required to directly control the many degrees of freedom for physical skills such as locomotion, object manipulation, gesture, and facial expression is overwhelming for the user. Hence, the robot would need to take high-level direction from the human, and be responsible for the performance of these physical and expressive abilities. To do so, the robot avatar would not need to have emotions itself, but it would need to be able to perceive and recognize the affective and linguistic intent of the user's message, and possess the expressive ability to faithfully express and convey this message to others.

Robot as partner. In this last paradigm, a person collaborates with the robot in a social manner as a person would with a teammate (Grosz, 1996). Robots that interact with people as capable partners need to possess social and emotional intelligence so that they can respond to and interact with people appropriately. Consider a robot that cares for an elderly person should be able to respond appropriately in times when the patient is showing signs of distress or anxiety. It should be persuasive in ways that are sensitive to the person, such as helping to remind them when to take medication, without being annoying or upsetting. It would need to know when to contact a health professional when necessary. Yet so many current technologies (animated agents, computers, etc.) interact with us in a manner characteristic of socially or emotionally impaired people. In the best cases they know what to do, but often lack the social-emotional intelligence to do it in an appropriate manner. As a result, they frustrate us and we quickly dismiss them even though they can be useful. Given that many exciting applications for autonomous robots in the future place them in a long-term relationship with people, robot designers need to address these issues or people will not accept robots into their daily lives.

4 Why Social/Sociable Robots?

In order to interact with others (whether it is a device, a robot, or even another person) it is essential to have a good conceptual model for how the other operates (Norman, 2001). With such a model, it is possible to explain and predict what the other is about to do, its reasons for doing it, and how to elicit a desired behavior from it. The design of a technological artifact, whether it is a robot, a computer, or a teapot, can help a person form this model by "projecting a image of its operation," either through visual cues or continual feedback (Norman, 2001). Hence, there is a very practical side to developing robots that can effectively convey and communicate their internal state to people for cooperative tasks, even when the style of interaction is not social.

For many autonomous robot applications, however, people will most likely use a social model to interact with robots in anthropomorphic terms. Humans are a profoundly social species. Our social-emotional intelligence is a useful and powerful means for understanding the behavior of, and for interacting with, some of the most complex entities in our world—people and other living creatures (Dennett, 1987). Faced with non-living things of sufficient complexity (i.e., when the observable behavior is not easily understood in terms of physics and its underlying mechanisms), we often apply a social model to explain, understand, and predict their behavior — attributing mental states (i.e., intents, beliefs, feelings, desires, etc.) to understand it (Reeves & Nass, 1996; Premack & Premack, 1995). Right or wrong, people rely on social models (or fluidly switch between using a social model with other mental models) to make complex behavior more familiar, understandable and intuitive. We do this because it is enjoyable for us, and it is often surprisingly quite useful (Dennett, 1987).

From a design perspective, the emotion system would implement much of the style and personality of a robot, encoding and conveying its attitudes and behavioral inclinations toward the events it encounters. Designing robots with personality may help provide people with a good mental model for them. According to Norman (2001), personality is a powerful design tool for helping people form a conceptual model that channels beliefs, behavior, and intensions in a cohesive and consistent set of behaviors. The parameters of the personality must fall within recognizable human (or animal) norms, however, otherwise the robot may appear mentally ill or completely alien. The robot's personality must be designed such that its behavior is understandable and predictable to people. Natural behavior can be a useful guide in this respect.

This raises an important question: to what extent does the robot's design support the social model? Simply stated, does applying a social mental model to understand and interact with the robot actually work? Many early examples of "social robots" (i.e. robot toys or entertainment robots) only project the surface appearance of possessing social and emotional intelligence. This may be acceptable for a sufficiently structured scenario (such as theme park entertainment, etc.) where the environment and the audience's interaction with the robot are highly constrained. However, as the complexity of the environment and the interaction scenario increases, the social sophistication of the robot will clearly have to scale accordingly. Once the robot's behavior fails to support the social model a person has for it, the usefulness of the model breaks down. Ideally, the robot's observable behavior will continue to adhere to a person's social expectations of it during natural interactions in the full complexity of the human environment. We argue that it will not be possible to achieve this degree of scalability without tackling the (hard) problem of developing "deep" architectures for socially and emotionally intelligent robots (see the chapter by Sloman in this volume).

5 Designing Sociable Robots

The *Sociable Robots Project* develops expressive anthropomorphic robots to explore scientific questions and to address engineering challenges of building socially and emotionally intelligent robots. Their social and emotive qualities are integrated deep into the core of their design, and serve not only to "lubricate" the interface between itself and its human interlocutor, but also play a pragmatic role in promoting survival, self maintenance, learning, decision making, attention, and more (Breazeal, 2002*a*, 2003*c*,*d*). Hence, social and affective interactions with people are valued not just at the interface, but at a pragmatic and functional level for the robot as well.

Humans, however, are the most socially and emotionally advanced of all species. As one might imagine, an autonomous anthropomorphic robot that could interpret, respond, and deliver human-style social and affective cues is quite a sophisticated machine. We have explored the simplest kind of human-style social interaction (guided and inspired by that which occurs between a human infant with its caregiver) and have used this as a metaphor for building a sociable robot, called Kismet (shown in Figure 1). The robot has been designed to support several social and emotive skills and mechanisms that are outlined in the rest of this chapter. Kismet is able to use these capabilities to enter into rich, flexible, dynamic interactions with people that are physical, affective, and social.

5.1 Expression of Affective State

Kismet can communicate its emotive state and other social cues to a human through facial expressions (Breazeal, 2003*a*), body posture, gaze direction (Breazeal et al., 2001), and quality of voice (Breazeal, 2003*b*). We do not have sufficient space to explain in detail how each of these are implemented, but they all contribute to the readability of Kismet's expression and its ability to communicate its internal state to a human in a natural and intuitive way.

We have found that the scientific basis for how emotion correlates to facial expression or vocal expression is very useful in mapping Kismet's emotive states to its face actuators (Breazeal, 2003*a*), and to its articulatory-based speech synthesizer (Breazeal, 2003*b*). In human-robot interaction studies (Breazeal, 2002*b*), we have found that these expressive cues are effective in regulating affective/intersubjective interactions (Trevarthen, 1979) and proto-dialogs (Tronick et al., 1979) between the human and the robot that resemble their natural correlates during infant-caregiver exchanges.

With respect to communicating emotion through the face, psychologists of the componential theory of facial expression posit that these expressions have a systematic, coherent, and meaningful structure that can be mapped to affective dimensions that span the relationship between different emotions (Smith & Scott, 1997). Some of the individual features of expression have inherent signal value. The raised brows, for instance, convey attentional activity for both the expression of fear and surprise. By considering the individual facial action components that contribute to the overall facial display, it is possible to infer much about the underlying properties of the emotion being expressed. This promotes a signaling system that is robust, flexible, and resilient. It allows for the mixing of these components to convey a wide range of affective messages, instead of being restricted to a fixed pattern for each emotion.

Inspired by this theory, Kismet's facial expressions are generated using an interpolation-based technique over a three-dimensional affect space (see Figure 2). The three dimensions correspond to arousal (high/low), valence (good/bad), and stance (advance/withdraw) — the same three attributes that are used to affectively assess the myriad of environmental and internal factors that contribute to Kismet's overall affective state (described in section 8.3). There are nine *basis postures* that collectively span this space of emotive expressions.

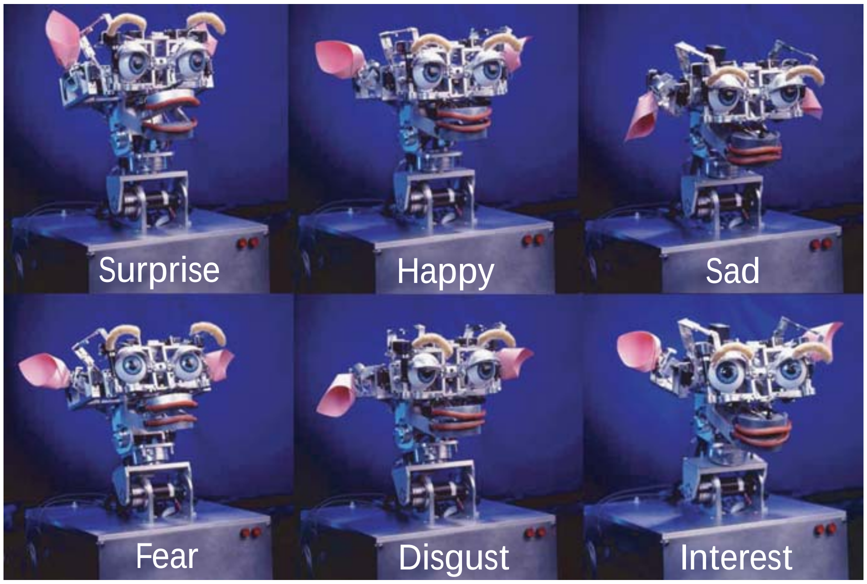

Figure 1: A sample of Kismet's facial expressions for basic emotions (see text). Kismet is about 1.5 times the size of an adult human head and has a total of 21 degrees of freedom (DoF). The robot perceives a variety of natural social cues from visual and auditory channels. Kismet has four cameras to visually perceive its environment: one behind each eye for post-attentive visual processing (such as face detection); one between the eyes to provide a wide peripheral view (to track bright colors, skin tone, and movement); and one in the "nose" that is used in stereo with the peripheral view camera to estimate the distance to targeted objects. A human wears a lavalier microphone to speak to the robot (see section 8.2.1). Images courtesy of Sam Ogden, copyright 2000.

The current affective state of the robot (as defined by the net values of arousal, valence, and stance) occupies a single point in this space at a time. As the robot's affective state changes, this point moves around this space and the robot's facial expression changes to mirror this. As positive valence increases, Kismet's lips turn upward, the mouth opens, and the eyebrows relax. However, as valence decreases, the brows furrow, the jaw closes, and the lips turn downward. Along the arousal dimension, the ears perk, the eyes widen, and the mouth opens as arousal increases. Along the stance dimension, the robot leans toward (increasing) or away (decreasing) from the stimulus. The expressions become more intense as the affect state moves to more extreme values in the affect space.

Hence, Kismet's face functions as a window by which a person can view the robot's underlying affective state. This transparency plays an important role in providing the human with the necessary feedback to understand and predict the robot's behavior when coupled with biologically-inspired emotive responses.

6 Architectural Overview

Inspired by models of intelligence in natural systems, the design of our architecture features both a cognitive system and an emotion system (see Figure 3).

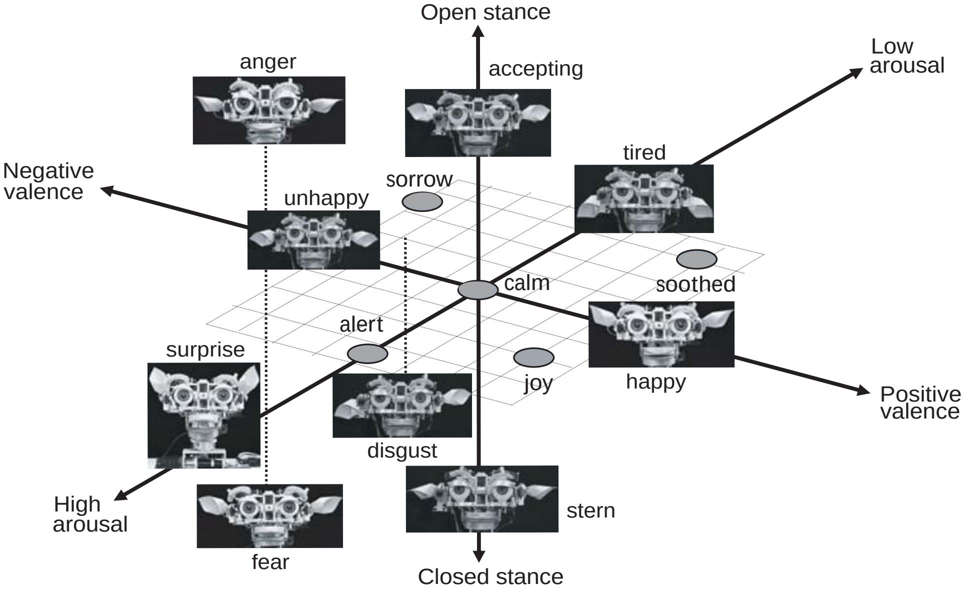

Figure 2: This diagram illustrates where the basis postures are located in Kismet's 3D affect space. The dimensions correspond to arousal (high or low), valence (good or bad), and stance (advance or withdraw). This space is used to generate Kismet's facial expressions based on the robot's overall affective assessment of the current situation. A sampling of where specific emotion categories map onto this space are shown as well.

Both operate in parallel and are deeply intertwined to foster appropriately adaptive functioning of the robot in the environment as it interacts with people.

It is common for biologically inspired architectures to be constructed from a network of interacting elements (e.g., subsumption architecture (Brooks, 1986), neural networks (McCulloch & Pitts, 1943), or agent architectures (Minsky, 1988)). Ours is implemented as an agent architecture where each computational element is conceptualized as a *specialist* (Minsky, 1988). Hence, each drive, behavior, perceptual releaser, motor, and emotion-related process is modeled as a different type of specialist that is specifically tailored for its role in the overall system architecture.

Each specialist receives messages from those connected to its inputs, performs some sort of specific computation based on these messages, and then sends the results to those elements connected to its outputs. Its *activation level*, A, is computed by the equation $A = (\sum_{j=1}^{n} w_j i_j) + b$ for integer values of inputs i_j, weights w_j, and bias b over the number of inputs n. The weights can be either positive or negative; a positive weight corresponds to an excitatory connection, and a negative weight corresponds to an inhibitory connection. Each process is responsible for computing its own activation level. The process is active when its activation level exceeds an *activation threshold*, T. When active, the process can send activation energy to other nodes to favor their activation. It may also perform some special computation, send output messages to connected processes, and/or express itself through motor acts by sending outputs to actuators. Hence, although the specialists differ in function, they all

	Facial Action							
Meaning	Eyebrow Frown	Raise Eyebrows	Raise upper Eyelid	Raise Lower Eyelid	Up Turn Lip Corners	Open Mouth	Tighten Mouth	Raise Chin
Pleasantness	⬇				⬆	⬆	⇓	⬇
Goal Obstacle/Discrepancy	⬆							
Anticipated Effort	⇧							
Attentional Activity		⬆	⬆					
Certainty		⇩		⇑		⇧		
Novelty		⇧	⇧					
Personal Agency/Control		⇩	⇩			⇩		

Table 1: A possible mapping of facial movements to affective dimensions proposed by Smith & Scott (1997). An up arrow indicates that the facial action is hypothesized to increase with increasing levels of the affective meaning dimension. A down arrow indicates that the facial action increases as the affective meaning dimension decreases. For instance, the lip corners turn upwards as "pleasantness" increases, and lower with increasing "unpleasantness."

follow this basic activation scheme.

Units are connected to form networks of interacting processes that allow for more complex computation. This involves connecting the output(s) of one unit to the input(s) of other unit(s). When a unit is active, besides passing messages to the units connected to it, it can also pass some of its activation energy. This is called *spreading activation* and is a mechanism by which units can influence the activation or suppression of other units. This mechanism was originally conceptualized by Lorenz (1973) in his *hydraulic model* of behavior. Minsky (1988) uses a similar scheme in his ideas of memory formation using *K-lines*. Popular architectures of Brooks (1986) and Maes (1991) are similar in spirit. However, ours is heavily inspired by ethological models and hence is most similar to that of Blumberg et al. (1996).

Ethology, comparative psychology, and neuroscience have shown that observable behavior is influenced by internal factors (i.e., motivations, past experience, etc.) and by external factors (i.e., perception). This demands that different types of systems be able to communicate and influence each other despite their different functions and modes of computation. This has led ethologists such as McFarland & Bosser (1993) and Lorenz (1973) to propose that there must be a *common currency*, shared by perceptual, motivational, and behavioral systems. Furthermore, as the system becomes more complex, it is possible that some components may conflict with others (e.g., competing for shared resources such as motor or perceptual abilities of the creature). In this case, the computational processes must have some means for competing for expression.

Based upon the use of common currency, Kismet's architecture is implemented as a *value-based system*. This simply means that each process computes numeric values (in a common currency) from its inputs. These values are passed as messages (or activation energy) throughout the network, either within a system or between systems. Conceptually, the magnitude of the value

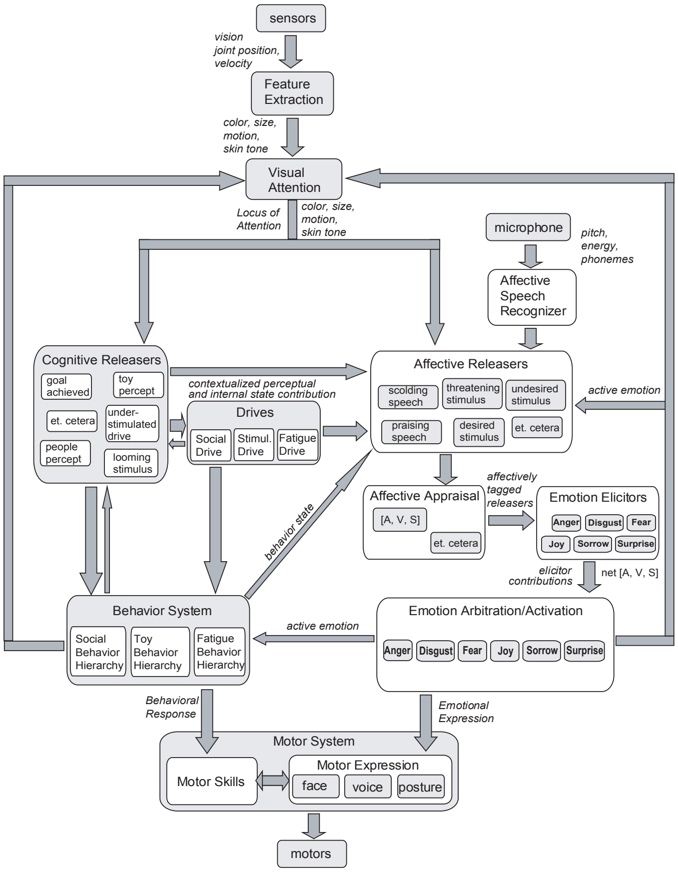

Figure 3: An architectural overview showing the tight integration of the cognitive system (shown in light grey), the emotion system (shown in white), and the motor system. The cognitive system is comprised of perception, attention, drives, and behavior systems. The emotion system is comprised of affective releasers, appraisals, elicitors, and "gateway" processes that orchestrate emotive responses. See text.

represents the strength of the contribution in influencing other processes. Using a value-based approach has the nice effect of allowing influences to be graded in intensity, instead of simply being on or off. Other processes compute their relevance based on the incoming activation energies or messages, and use their computed activation level to compete with others for exerting influence upon the other systems.

7 Overview of the Cognitive System

The cognitive system is responsible for perceiving and interpreting events, and for arbitrating among the robot's goal-achieving behaviors to address competing motivations. There are two kinds of motivations modeled in Kismet. The *drives* reside in the cognitive system and are modeled as a homeostatic processes that represent the robot's "health" related goals. The emotion system also motivates behavior as described in section 8.

The computational sub-systems and mechanisms that comprise the cognitive system work in concert to decide which behavior to activate, at what time, and for how long to service the appropriate objective. Overall, the robot's behavior must exhibit an appropriate degree of relevance, persistence, flexibility, and robustness. To achieve this, we based the design of the cognitive system on ethological models of animal behavior (Gould, 1982). In section 9, we discuss how Kismet's emotion-inspired mechanisms further improve upon the basic decision-making functionality provided by the cognitive system.

7.1 Perceptual Elicitors

Sensory inputs arising from the environment are sent to the perceptual system where key features are extracted from the robot's sensors (cameras, microphones, etc.). These features are fed into an associated *releaser* process. Each releaser can be thought of as a simple perceptual elicitor of behavior that combines lower-level features into behaviorally significant perceptual categories. For instance, the visual features of `color`, `size`, `motion`, `proximity` are integrated to form a `toy` percept. Other releasers are designed to characterize important internal events, such as the urgency to tend to a particular motive. There are many different releasers designed for Kismet (too many to list here), each signals a particular event or object of behavioral significance. If the input features are present and of sufficient intensity, the activation level of the releaser process rises above threshold, signifying that the conditions specified by that releaser hold. Given this, its output is passed to its corresponding behavior process in the behavior system, thereby preferentially contributing to that behavior's activation. Note that Kismet is not a stimulus-response system, given that internal factors (i.e. motivations as defined by drives and emotions) also contribute to the robot's behavior activation.

7.2 Cognitive Motivations: Drives

Within the cognitive system, Kismet's *drives* implement autopoetic-related processes for satisfying the robot's "health" related and time-varying goals (Maturana & Varela, 1980). Analogous to the motivations of "thrist," "hunger," and "fatigue" for an animal, Kismet's drives motivate it to get the right amount of the desired kind of stimulation in a timely manner. Kismet's drives correspond to a "need" to interact with people (the `social-drive`), to be stimulated by toys (the `stimulation-drive`), and to occasionally rest (the `fatigue-drive`).

In living creatures, these autopoetic processes are innate and are directly tied to the animal's physiology (see the paper by Kelley in this volume). The design of each drive in Kismet is heavily inspired by ethological views of the analogous process in animals where their change in intensity reflects the ongoing "needs" of the creature and the urgency for tending to them.

Each drive acts to maintain a level of intensity within a bounded range, neither too much nor too little, as defined by a desired operational point and acceptable bounds of operation around that point (called the homeostatic regime). A drive remains in its homeostatic regime when it is encountering its satiatory stimulus of appropriate intensity. Given no satiatory stimulation, a drive will tend to increase in (positive) intensity. The degree to which each drive is satiated in a timely fashion contributes to the robot's overall measure of its "well being."

This information is also assigned affective value by the emotion system (described in section 8.3). For example, negative value is assigned when the robot's needs are not being met properly, and positive value is assigned when they are. This is a rough analogy to the discussion of rewards and punishers associated with homeostatic need states in the chapter by Rolls. Hence, the affective contribution of the drives do not directly evoke emotive responses, but they do bias the robot's net affective state. In this sense, the drives contribute to the robot's "mood" over time, which makes the corresponding emotive responses easier to elicit.

Drives shape the internal agenda of the robot (in concert with perceptual and emotive factors) and play an important role in determining which behavior to next engage. To keep its activation level within the homeostatic regime, each drive can preferentially spread activation to behaviors at the top level of the behavior hierarchy that help to restore that drive (described in detail in the following section).

Behaviors, in turn, encode specific task-achieving goals that serve to maintain the robot's internal state (as defined by the state of the drives and emotions). To remain in balance (near the center of the spectrum), it is not sufficient that the satiatory stimulus merely be present; it must also be of a good quality. For instance, in the absence of the satiatory stimulus (or if the intensity is too low), a drive increases in intensity to the positive end of the spectrum and preferentially biases the activation of those behaviors that serve to seek out that stimulus. In addition, the affective contribution of the drive (negative valence and low arousal) contributes to a net affective state that makes it easier for the sorrow emotive response to become active. Sorrow represents a different strategy to help the robot come into contact with a desired stimulus by signaling to people that the robot needs attention.

Alternatively, if the satiatory stimulus is too intense (e.g., a visual stimulus is moving too fast or is too close to the robot's face), a drive tends toward the extreme negative end of the spectrum. In this circumstance, the drive biases the activation of avoidance behaviors to limit the robot's exposure to the intense stimulus. Also, the affective contribution of the drive (negative valence, high arousal) contributes to a net affective state that makes it easier for Kismet's fear response to become active. Once active, the fearful expression on Kismet's face signals to people to back off a bit.

Hence, the drives work in concert with behaviors and contribute to an affective state that helps Kismet keep its level of interaction with the world and people in balance, not too much nor too little.

7.3 Behavior Arbitration as Decision Making

Within the behavior system, the behavior processes are organized into loosely layered, heterogeneous hierarchies of behavior groups (see Figure 4), much in the spirit of those ethological models proposed by Tinbergen (1951) and Lorenz (1973). Implicit in this model is that a decision is being made among several alternatives at every level of the hierarchy—of which one is chosen.

Functional Groups. The decisions are very general at the top of the hierarchy (which drive to satiate) and become increasingly more specific as one moves down the hierarchy. At the topmost level, behaviors are organized into competing functional groups (i.e., the primary branches of the hierarchy—there are

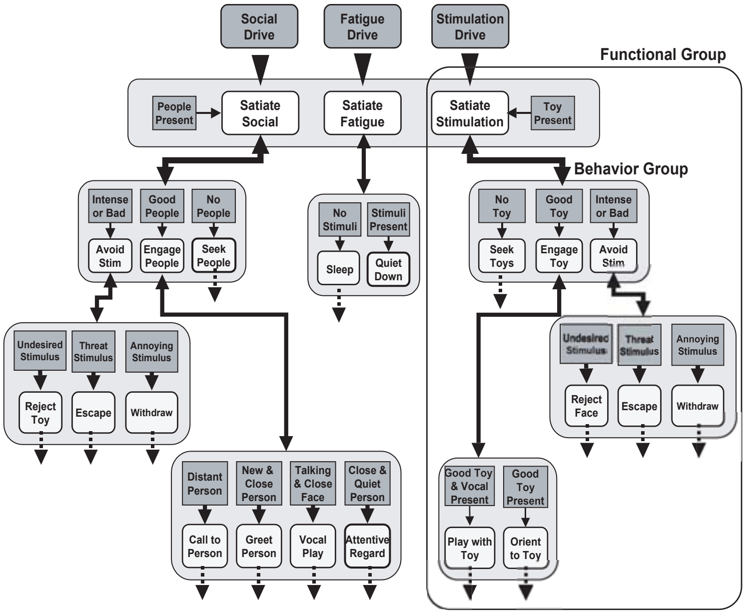

Figure 4: Schematic of Kismet's behavior hierarchy that resides within the behavior system of Kismet's cognitive-affective architecture. The affective factors for each behavior (both inputs and outputs) are not shown in this diagram. Dashed arrows represent connections to the motor system. Input from the drives feed into the top level of the hierarchy, as shown by the dark gray boxes with rounded edges. Functional groups are represented as the major branches of the hierarchy. For instance, the functional group for satiating Kismet's drive to interact with toys is highlighted. Behavior groups are shown as light gray boxes with rounded edges, containing competing behaviors (white boxes with rounded edges) and their perceptual elicitors (dark gray boxes). See text.

three in Kismet). Each functional group is responsible for maintaining one of the three homeostatic functions and only one functional group can be active at a time. This property is inspired by animal behavior where an animal engages in one class of activities at a time: either foraging for food, sleeping, defending territory, mating, etc. Thus, the influence of the drives is strongest at the top level of the hierarchy, biasing which functional group should be active.

Behavior Groups. Each functional group consists of an organized hierarchy of behavior groups that are akin to Tinbergen's behavioral centers (Tinbergen, 1951). At each level within a functional group hierarchy, each behavior group represents a competing strategy (a collection of behaviors) for satisfying the goal of its parent behavior.

Behaviors. Each behavior within a behavior group is viewed as a task-achieving entity whose particular goal contributes to the strategy of its behavior group. Each behavior process within a group competes with the others in a winner-take-all fashion for expression based on its measure of relevance to the current situation. A behavior's measure of relevance takes into account several factors including: the perceived environment through its releaser inputs; internal motives through its drive and emotion inputs; and internally computed progress-measures such as its level of interest (how long this behavior has been active) and its level of frustration (a measure of progress toward its goal over time). When active, a behavior coordinates sensori-motor patterns to achieve a particular task such as search behaviors, approach behaviors, avoidance behaviors, and interaction behaviors. The perceptual, homeostatic, and affective factors that contribute to behavioral relevance allow the robot to act in a manner that is persistent (such as trying new strategies to attain a blocked goal) but also suitably opportunistic (described in more detail in section 9).

Therefore, the observed behavior of the robot is the result of competition at the functional, strategic, and task levels. At the behavioral category level, the functional groups compete to determine which "need" is to be met (for Kismet, this corresponds to socializing, playing, or sleeping). At the strategy level, behavior groups of the winning functional group compete for expression. For instance, two of the behavior groups at the second level contain several strategies for acquiring a desired stimulus of an appropriate intensity: `seek` or acquire the desired stimulus if it is too weak or not present; `avoid` or reduce the intensity of the stimulus if it is too overwhelming; or `engage` the stimulus if it is of appropriate intensity. Finally, on the task level, the behaviors of the winning behavior group compete for expression to determine which sub-goal the robot pursues (i.e., ways of acquiring the desired stimulus, ways of reducing its intensity if it overwhelming, and ways of engaging the stimulus when it is appropriate). As one moves down in depth, the behaviors serve to more finely tune the relation between the robot and its environment, and in particular, the relation between the robot and the human (Breazeal, 2002*a*).

8 Overview of the Emotion System

The emotion system is responsible for perceiving and recognizing internal and external events with affective value, assessing and signaling this value to other systems, regulating and biasing the cognitive system to promote appropriate and flexible decision making, and communicating the robot's internal state. Kismet communicates its emotive state in a transparent and familiar manner through facial expression, body posture and tone of voice. This makes the robot's behavior more predictable and understandable to the person who is interacting with it. These expressive behaviors allow Kismet to socially regulate people's behavior toward it in a natural way that is beneficial to the robot.

Thus, in concert with the robot's cognitive system, the emotion system is designed to be a flexible and complementary system that mediates between environmental, social, and internal events to elicit an adaptive behavioral response that serves either social or self-maintenance functions. In animals and people, each specific emotion motivates and coordinates cognitive systems and patterns of behavior responses to facilitate development, adaptation, and coping in a particular way. The remainder of section 8 outlines how Kismet's emotional responses are implemented. Broadly speaking, each emotive response consists of:

- A precipitating event (section 8.2).
- An affective appraisal of that event (section 8.3).

- A characteristic display that can be expressed through facial expression, vocal quality, or body posture (section 5.1).
- Modulation of the cognitive and motor systems to motivate a behavioral response (sections 8.4 and 9).

According to Rolls (see his chapter in this volume), emotions can be pragmatically defined as states elicited by rewards and punishments where a reward is something for which an animal will work and a punishment is something it will work to escape or avoid. Kismet's affective appraisal process involves assessing whether something is rewarding or punishing and tagging it with a value that reflects its expected benefit or harm to the robot. The emotion system combines the myriad of affectively tagged inputs to compute the net affective state of the robot. Similar to Rolls, the affective tags serve as the common currency for the inputs to the response selection mechanism. Thus, they are used to determine the most relevant emotive response for the given situation. Once a particular emotive response is active, Kismet engages in a process of *behavioral homeostasis*, where the active emotive response maintains behavioral activity in its particular manner through external and internal feedback until the correct relation of robot to environment is established (Plutchik, 1991).

Kismet's desired internal and external relationship is comprised of two factors. First, whether the robot's homeostatic needs are being met in a timely manner. Second, whether its net affective state corresponds to a mildly positive, aroused, and open state. When Kismet's internal state diverges from this desired internal relationship, the robot will work to restore the balance—to acquire desired stimuli, to avoid undesired stimuli, and to escape dangerous stimuli. Each emotive response carries this out in a distinct fashion by interacting with the cognitive system to evoke a characteristic behavioral pattern and to socially cue others as to whether the interaction is appropriate or not (and how they might respond to correct the problem). A detailed description of the implementation of each emotive response can be found in Breazeal (2002*a*).

8.1 Functions of Basic Emotions

The organization and operation of Kismet's emotion system is strongly inspired by various theories of *basic emotions* in humans and animals (Ekman, 1992). These few select emotions are endowed by evolution because of their proven ability to facilitate adaptive responses that promote a creature's daily survival in a complex and often hostile environment. Basic emotions are termed "independent" emotions because their emergence does not require or reduce to cognitive processes (Ackerman et al., 1998). As shown in Table 2, a number of basic emotive responses have been implemented on Kismet. This table is derived from the cross-species and social functions hypothesized by Izard & Ackerman (2000) and Plutchik (1991).

Anger. In living systems, anger serves to mobilize and sustain energy and vigorous motor activity at high levels (Tomkins, 1963). It is often elicited when progress toward a goal is hindered or blocked. Similarly, in Kismet, a frustrated state (increasing in intensity to anger) arises when progress toward the current goal is slow. This mobilizes the robot to try alternate strategies.

Antecedent Conditions	*Emotion*	*Behavior*	*Function*
Delay, difficulty in achieving goal of active behavior	anger, frustration	Display agitation, energize	Show displeasure to modify human's behavior. Try new behavior to surmount blocked goal.
Presence of an undesired stimulus	disgust	Withdraw	Signal rejection of presented stimulus
Presence of a threatening or overwhelming stim.	fear, distress	Display fear, escape	Move away from a potentially dangerous stimuli
Prolonged presence of a desired stimulus	calm	Engage	Continued interaction with a desired stimulus
Success in achieving goal of active behavior, or praise	joy	Display pleasure	Reallocate resources to the next relevant behavior
Prolonged absence of a desired stim., or scolding	sorrow	Display sorrow	Evoke sympathy and attention from human.
A sudden, close stimulus	suprise	Startle response	Alert
Appearance of a desired stim.	interest	Orient, explore	Attend to new, salient object. Engage
Absense of stimulus	boredom	Seek	Explore environment for desired stimulus

Table 2: Summary of the antecedents and behavioral responses that comprise Kismet's emotive responses. The antecedents refer to the eliciting perceptual conditions for each emotion. The behavior column denotes the observable response that becomes active with the emotion. For some, this is simply a facial expression. For others, it is a behavior such as `escape`. The column to the right describes the function each emotive response serves Kismet.

Disgust. Tomkins (1963) describes disgust as a reaction to unwanted intimacy with a repellent entity. Generally speaking, disgust is manifested as a distancing from some object, event, or situation, and can be characterized as rejection (Izard, 1997). It is in this sense that disgust is incorporated into Kismet's emotive repertoire. Kismet's disgust response signals rejection of an unwanted stimulus.

Fear. The unique function of fear is to motivate avoidance or escape from a dangerous situation. For Kismet, the fear response protects it from possible harm when faced with a threatening stimulus that could cause damage (e.g., large stimuli moving fast and close to the robot's face). Kismet's fearful expression is an communicative cue that signals to a person that they should back off a bit (Breazeal & Scassellati, 2000). If they persist, the robot will evoke a protective `escape` response—e.g., close its eyes and turn its head away from the offending stimulus.

Joy. The emotion of joy is believed to heighten openness to experience. It often arises upon the success of achieving a goal or in the pleasure of mastery, exhibited even by very young children (Meltzoff & Moore, 1997). In humans, openness in social situations contributes to affiliative behavior and the strengthening of social bonds. The expression of joy operates as a universally recognizable signal of readiness for friendly interaction. For Kismet, it serves a social function to encourage people to interact with it. It also arises when the robot has achieved a pursued goal, accompanied by a reallocation of cognitive/behavioral resources to the next relevant task.

Sorrow. Izard argues that sadness is unique in its capacity to slow the cognitive and motor systems. Tomkins (1963) suggests that slowing down enables one to reflect upon a disappointing performance to gain a new perspective that will help improve future performance. Sadness can also strengthen social bonds. The expression of sorrow communicates to others that one is in trouble and increases the likelihood that the others will feel sympathy and lend assistance (Moore et al., 1984). Similarly, Kismet's expression of sorrow serves a communicative function that encourages people to pay attention to it and to try to cheer it up (Breazeal, 2002*b*). In human-robot interaction studies, we have found the Kismet's expression elicits sympathetic responses in people (Breazeal, 2002*b*).

Surprise. Children show surprise when there are violations of expected events, or as a response to discovery such as an "Aha!" experience. Hence cognitive processes play an important role in the emergence of this early emotion. Given that simple cognitions elicit surprise, some emotion theorists do not consider surprise to be a basic emotion even though it appears early in infancy (around 6 months of age). For Kismet, surprise is a startle response that is elicited by a sudden and unexpected event, such as a quickly looming stimulus.

Interest. In animals and people, interest motivates exploration, learning, and creativity. It mobilizes the creature for engagement and interaction. It serves as a mechanism of selective attention that keeps the creature's attention focused on a particular object, person, or situation, and away from other distractions that impinge upon its senses. For Kismet it serves a similar function with respect to focusing attention and motivating exploration and interaction.

Boredom. In Kismet, boredom is treated as a basic emotion that arises when the robot is not stimulated for a while. Over the long term, this prolonged absence will elicit sorrow. In the shorter term, however, boredom motivates searching behaviors similar to interest—however, its function is to come into contact with a desired stimulus, rather than to engage one that is already present.

8.2 Affective Releasers

The *affective releasers* assess the value of perceptual inputs arising from the environment. They are similar to the perceptual releasers of the cognitive system, but rather than only being a perceptual interpretation of stimuli into objects and events, they are also cognitively appraised in relation to the motivational state of the robot and its current goals. Beyond simple perceptual features, the affective releasers go through a more detailed cognitive appraisal to judge their expected benefit to the robot: the quality of the stimulus (e.g., the intensity is too low, too high, or just right), or whether it is desired or not (e.g., it relates to the active goals or motivations). For instance, if the `stimulation-drive` is being tended to and a nearby toy is moving neither too fast nor too close to the robot, then the `desired-toy` releaser is active. However, if the `social-drive` is being tended to instead, then the `undesired-toy` releaser is active. If the toy has an aggressive motion (i.e., too close and moving too fast), then the `threatening-toy` releaser is active. This evaluation is converted into an activation level for that affective releaser. If the activation level is above threshold, then its output is passed to the affective appraisal stage where it can influence the net affective state and emotive response of the robot.

8.2.1 Recognition of communicated affect

Objects that Kismet interacts with can have affective value, such as a toy that is moving in a threatening manner. However, people can communicate affect directly to Kismet through tone of voice. Developmental psycholinguists can tell us quite a lot about how preverbal infants achieve this. Based on a series of cross-linguistic analyses, there appear to be at least four different pitch contours that pre-linguistic infants can recognize affectively (i.e., approval, prohibition, comfort, and attentional bids), each associated with a different emotional state (Fernald, 1989). Characteristic prosody curves for each are shown in figure 5.

Inspired by these theories, we have implemented a recognizer for distinguishing these four distinct prosodic patterns from Kismet-directed speech. The implemented classifier consists of several mini-classifiers executing in stages (see Figure 6). A very detailed presentation of the recognizer and its performance assessment can be found in Breazeal & Aryananda (2002).

Based on our recordings, the preprocessed pitch contours from the training set resemble Fernald's prototypical prosodic contours for approval, attention, prohibition, comfort/soothing, and neutral. Hence, we used Fernald's insights to select those features that would prove useful in distinguishing these five classes. For the first classifier stage, global pitch and energy features (i.e., pitch mean and energy variance) partitioned the samples into useful intermediate classes (see Figure 7).

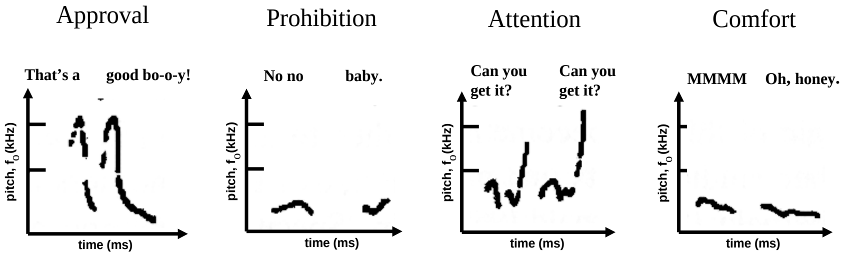

Figure 5: Fernald's prototypical contours for approval, prohibition, attention, and soothing. It is argued that they are well-matched to saliency measures hardwired into an infant's auditory processing system (Fernald, 1989).

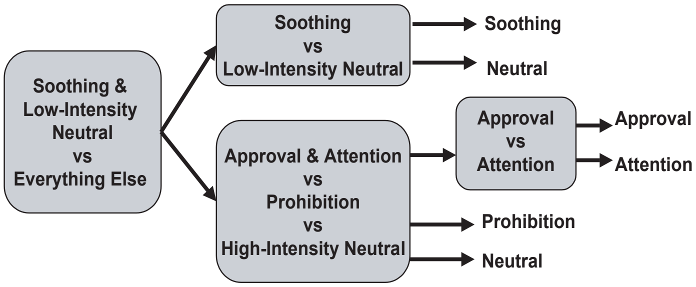

Figure 6: The spoken affective intent recognizer. Each stage is a mini-classifier that uses acoustic features identified by Fernald (1989) to recognize each of the four affective intents.

For instance, the prohibition samples are clustered in the low pitch mean and high energy variance region. The approval and attention classes form a cluster at the high pitch mean and high energy variance region. The soothing samples are clustered in the low pitch mean and low energy variance region. Finally, the neutral samples have low pitch mean, but are divided into two regions in terms of their energy variance values. The structure of each of the mini-classifiers follows logically from these observations. Table 3 shows the resulting classification performance using a Gaussian mixture model, updated with the EM algorithm, to represent the distribution of data. The output of each affective intent classifier is treated as an affective releaser and sent with the others to the affective appraisal phase.

8.3 Affective Appraisal

In Kismet's implementation, there is an explicit assessment phase for each active releaser, of which there are a number of factors that contribute to the assessment made. The assessment consists of labeling the releaser with affective tags, a mechanism inspired by Damasio's *somatic marker hypothesis*, where incoming

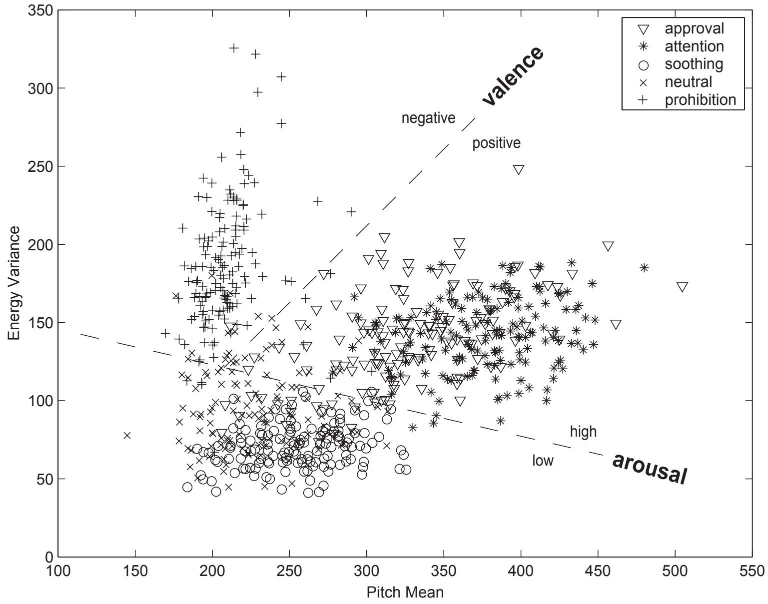

Figure 7: Feature space of all five classes with respect to energy variance and pitch mean. There are three distinguishable clusters (roughly partitioned by arousal and valence) for prohibition, soothing and neutral, and approval and attention.

perceptual, behavioral, or motivational information is "tagged" with affective information (Damasio, 1994). The tagged value reflects whether the releaser is expected to be rewarding or punishing to the robot.

For example, there are three classes of tags used within Kismet to affectively characterize a given releaser. Each tag has an associated intensity that scales its contribution to the overall affective state. The arousal tag, A, specifies how arousing this factor is to the emotional system. It very roughly corresponds to the activity of the autonomic nervous system. Positive values correspond to a high arousal stimulus whereas negative values correspond to a low arousal stimulus. The valence tag, V, specifies how favorable or unfavorable this percept is to the robot. Positive values correspond to a beneficial stimulus whereas negative values correspond to a stimulus that is not beneficial. The stance tag, S, specifies how approachable the percept is. Positive values correspond to advance whereas negative values correspond to retreat.

There are three factors that contribute to an appraisal of an active releaser. The first is the intensity of the stimulus, which generally maps to arousal. Threatening or very intense stimuli are tagged with high arousal. Absent or low intensity stimuli are tagged with low arousal. The second is the relevance of the stimulus to whether it addresses the current goals of the robot. This influences the valence and stance values. Stimuli that are relevant are desirable. They are tagged with positive valence and approaching stance. Stimuli that are not relevant are undesirable. They are tagged with negative arousal and withdrawing stance. The third factor is intrinsic pleasantness. Some stimuli are hardwired to influence the robot's affective state in a specific manner. For in-

Catgy	*Test Size*	*Appr*	*Attn*	*Prohib*	*Comft*	*Ntrl*	*% Corrct*
Appr	84	64	15	0	5	0	76.2
Attn	77	21	55	0	0	1	74.3
Prohib	80	0	1	78	0	1	97.5
Comft	68	0	0	0	55	13	80.9
Ntrl	62	3	4	0	3	52	83.9
All	371						81.9

Table 3: Overall classification performance evaluated using a new test set of 371 utterances from five adult female speakers ranging in age from 23 to 54 years old.

stance, praising speech is tagged with positive valence and slightly high arousal, whereas scolding speech is tagged with negative valence and low arousal, which tends to elicit as sorrowful response. In Kismet, there is a fixed mapping from each of these factors to how much they contribute to arousal, valence, or stance.

In addition to the perceptual contribution of the releasers, other internal factors can also influence the robot's emotive state. For instance, the drives contribute according to how well they are being satiated. The homeostatic regime is marked with positive valence and balanced arousal, contributing to a contented affective state. The under stimulated regime (large positive values) is marked with negative valence and low arousal, contributing to a bored affective state that can eventually decline to sorrow. The over stimulated regime (large negative values) is marked with negative valence and high arousal, contributing to an affective state of distress. Another factor is progress towards achieving the desired goal of the active behavior. Success in achieving a goal promotes joy and is tagged with positive valence. Prolonged delay in achieving a goal results in frustration and is tagged with negative valence and withdrawn stance. It is also possible for the active emotion to either contribute to or inhibit the activation of other emotions, making it difficult for a creature to be both happy and angry simultaneously, for instance.

Because there are potentially many different kinds of factors that modulate the robot's affective state (e.g., behaviors, motivations, perceptions), this tagging process converts the myriad of factors into a common currency that can be combined to determine the net affective state. For Kismet, the $[A, V, S]$ trio is the currency the emotion system uses to determine which emotional response should be active. In the current implementation, the values of the affective tags for the releasers are specified by the designer. These may be fixed constants, or linearly varying quantities.

8.4 Emotion Elicitors and Arbitration

All somatically marked inputs (e.g., releasers, the state of each drive, etc.) are passed to the emotion elicitors. There is an elicitor associated with each basic emotion "gateway" process (e.g., `anger`, `fear`, `disgust`, etc.). The elicitor determines the relevance of its emotive response based on the myriad of factors contributing to it. In a living creature, this might include neural factors, sensorimotor factors, motivational factors, and cognitive factors (Izard, 1993). Each elicitor computes the relevance of its affiliated emotion process and contributes to its activation. Each elicitor can thus be modeled as a process that computes its activation energy, $E_{emot}(i)$, for emotion, i, according to the function,

$$E_{emot}(i) = R_{emot}(i) + Dr_{emot}(i) + Em_{emot}^{excite}(i) - Em_{emot}^{inhibit}(i) + Bh_{emot}(i)$$

Given the following somatically marked factors: $R_{emot}(i)$ is the weighted contribution of the active releasers, $Dr_{emot}(i)$ is the weighted contribution of the active drive, $Em_{emot}^{excite}(i)$ is the weighted contribution of the other active emotions that excite this process, $Em_{emot}^{inhibit}(i)$ is the weighted contribution of the other active emotions that inhibit this process, and $Bh_{emot}(i)$ is the weighted contribution of the behavioral progress towards the current goal.

Each emotion "gateway" processes competes for control in a winner-take-all arbitration scheme based on their activation level. Although these processes are always active, their intensity must exceed a threshold level before they are expressed externally. The activation of each process is computed by the equation,

$$A_{emot}(i) = \sum_{i}(E_{emot}(i) + b_{emot}(i) + p_{emot}(i)) - \delta_t(i)$$

Where $E_{emot}(i)$ is the activation level computed by the affiliated emotive elicitor process described above, $b_{emot}(i)$ is a constant offset that can be used to make the emotion processes easier or harder to activate in relation to the activation threshold , and $p_{emot}(i)$ adds a level of persistence to the active emotion. This introduces a form of inertia so that different emotion processes don't rapidly switch back and forth. Finally, $\delta_t(i)$ is a decay term that restores an emotion to its bias value once the emotion becomes active.

When active, each emotion "gateway" process acts as a gateway that when "open" can spread activation to a number of different cognitive systems (i.e., behavior, attention, expression, etc.). As a result, the emotive state of the robot is distributed throughout the overall architecture; it does not reside in the gateway process itself.

Each emotion "gateway" process plays a distinct regulatory role, modulating the cognitive and expressive systems in a characteristic manner when active. In a process of behavioral homeostasis, the emotive response maintains activity through external and internal feedback until the correct relation of robot to environment is established (Plutchik, 1991). Concurrently, the affective state of the robot, as specified by the net $[A, V, S]$ of the active process, is sent to the expressive components of the motor system, causing a distinct facial expression, vocal quality, and body posture to arise. A detailed description of the implementation of each emotive response can be found in Breazeal (2002*a*).

9 Integrated Emotive Responses

In this section, we illustrate how Kismet's emotion system works in concert with its cognitive system to address its competing goals and motives given its limited resources and the ever-changing demands of interacting with people (Breazeal, 2002*a*). The emotion system achieves this by assessing and signaling the value of immediate events in order to appropriately regulate and bias the cognitive system to help focus attention, prioritize goals, and to pursue the current goal with an appropriate degree of persistence and opportunism. The emotive responses protect the robot from intense interactions that may be potentially harmful, and help the robot to sustain interactions that are beneficial for the robot. We highlight that the emotion system improves the performance of the robot over that provided by the cognitive system alone.

9.1 Communicative expression

Each emotive response entry of Table 2 is comprised of a goal-achieving behavioral component and an accompanying expressive display. For some of the emotive responses, the expressive display addresses both aspects when it serves a communicative function that is designed to elicit a desired behavioral response from the human that satisfies the robot's goal.

Kismet's expressive abilities successfully serve its goals when interacting with a person for two main reasons. First, we have found that people[1] enjoy playing with Kismet and want to sustain a pleasurable interaction with it (Breazeal & Scassellati, 2000; Breazeal, 2003*a*, 2002*a*). People find the robot to be lively, and to have an appealing personality and convincing social presence. This is a result of the way Kismet's emotion system is designed to interact with its cognitive system (as argued in section 4). Thus, both Kismet and the person have the shared goal of establishing and maintaining a beneficial interaction. The interaction is beneficial to the human if it is enjoyable, and it is beneficial to the robot if its motivations and goals are satisfied.

Second, Kismet's expressive behavior is effective because it is readily understandable and predictable to the person who interacts with it. This follows from the fact that Kismet's emotive responses are modeled after basic emotions which are universally understood by people (Ekman, 1992). As a result, people readily infer how they must adapt their behavior to get a desired response from Kismet—to keep it happy and interested and to avoid causing it distress.

For instance, Kismet exhibits sorrow upon the prolonged absence of a desired stimulus. This may occur if the robot has not been engaged with a toy for a long time. The sorrowful expression is intended to elicit attentive acts from the human analogous to how an infant's cries elicit nurturing responses from its caregiver. Kismet uses other expressive displays, such as "fear," to encourage people to slow down or back off a bit if they are crowding its cameras or moving too fast for the robot to perceive them. This allows the robot to tune the human's behavior so that it is appropriate for it. When the interaction is beneficial to Kismet, the robot conveys a state of interest and joy which encourages people to sustain the interaction. In a number of human-robot interaction studies with Kismet, we have found this to be quite effective as people find pleasure in cheering up the robot and keeping it engaged without being instructed to do so (Breazeal, 2003*a*).

9.2 Biasing attention

Kismet's level of interest improves the robot's attention, biasing it toward desired stimuli (e.g., those relevant to the current goal) and away from irrelevant stimuli. For instance, Kismet's exploratory responses include visual search for a desired stimulus and/or maintaining visual engagement of a relevant stimulus. Kismet's visual attention system directs the robot's gaze to the most salient object in its field of view, where the overall salience measure is a combination of the object's raw perceptual salience (e.g. size, motion, color) and its relevance to the current goal. It is important to note that Kismet's level of interest biases the robot to focus its attention on a goal-relevant stimulus that is beneficial to the robot, even when that object may have less perceptual salience over an-

---

[1] In a collaboration with sociologist Dr. Sherry Turkle, human subjects of different ages were brought in to participate in human-robot interaction studies with Kismet.

other "flashy" yet less goal-relevant stimulus. Without the influence of interest on Kismet's attention, the robot would end up looking at the "flashy" stimulus even if it has less behavioral benefit to the robot.

In addition, Kismet's disgust response allows it to reject and look away from an undesired stimulus. This directs the robot's gaze to another point in the visual field where it might find a more desirable object to attend. It also provides an expressive cue that tells the human that the robot wants to look at something else. The person often responds by trying to engage Kismet with a different toy, etc. This increases the robot's chances that it might be presented with a stimulus that is more appropriate to the robot's goal. We have found that people are quick to determine which stimulus the robot is after and readily present it to Kismet (Breazeal, 2003*a*, 2002*b*; Breazeal & Scassellati, 2000). This allows the robot to cooperate with the human to obtain a desired stimulus faster than it would if it had to discover one on its own.

9.3 Goal prioritization, persistence, and opportunism

Emotion and affect play an important role in helping Kismet to prioritize goals and to decide when to switch among them. They contribute to this process through a variety of mechanisms to make Kismet's goal pursuing behavior both flexible, opportunistic, and appropriately persistent.

Emotive influences. For instance, Kismet's fear response allows it to quickly switch from engagement behaviors to avoidance behaviors once an interaction becomes too intense or turns potentially harmful. This is an example of a rapid re-prioritization of goals. The fear response accomplishes this by effectively "hijacking" the behavior and motor systems to rapidly respond to the situation. For instance, the fear response may evoke Kismet's escape behavior, causing the robot to close its eyes and turn its head away from the offending stimulus.

Affective drive influences. In addition, affective signals arising from the drives bias which behaviors become active to satiate a particular motive. These affective influences contribute to activating behaviors that are the most relevant to the robot's "health" related needs. When the drives are reasonably well satiated, the perceptual contributions play the dominant role in determining which goals to pursue. Hence, the presence of a person will tend to elicit social behaviors and the presence of a toy will tend to elicit toy-directed behaviors. As a result, Kismet's behavior appears strongly opportunistic, taking advantage of whatever stimulus presents itself to the robot.

However, if a particular drive is not satiated for a while, its influence on behavior selection will grow in intensity. When this occurs, the robot becomes less opportunistic and grows more persistent about pursing those goals that are relevant to that particular drive. For instance, the robot's behavior becomes more "finicky" as it grows more prone to give a disgust response to stimuli that do not satiate that specific drive. The robot will also start to exhibit a stronger looking preference to stimuli that satiate that drive over those that do not. These aspects of persistent behavior continue until the drive is reasonably satiated again.

Affective behavior influences. Another class of affective responses influences arbitration between competing behavioral strategies to achieve the same goal. Delayed progress of a particular behavior results in a state of growing frustration, reflected by a stern expression on the robot's face. As Kismet grows more frustrated it lowers the activation level of the active behavior within the behavior group. This makes it more likely to switch to another behavior within the same group that could have a greater chance of achieving success of the current goal.

For instance, if Kismet's goal is to socialize with a person it will try to get a person to interact with it in a suitable manner (e.g., arousing but not too aggressive). If the perceptual system detects the presence of a person, but the person is ignoring Kismet, the robot will engage in behaviors to attract the person's attention. For instance, the robot's initial strategy might be to vocalize to the person to get his or her attention. If this strategy fails over a few attempts, the level of frustration associated with this behavior increases as its activation-level decreases. This gives other competing behaviors within the same behavior group a chance to win the competition and become active instead. For instance, the next active behavior strategy might be one where Kismet leans forward and wiggles its ears in an attention-grabbing display. If this also fails, eventually, the prolonged absence of social interaction will elicit sorrow, which encourages sympathy responses from people—a third strategy to get attention from people to satiate the `social-drive`.

10 Conclusion

In this chapter, we have explored the benefits that emotive and cognitive aspects bring to the design of autonomous robots that operate in complex and uncertain environments and perform in cooperation with people. Our examples highlight how Kismet's emotion system works intimately with its cognitive system to improve the overall performance of the robot. Although the cognitive system is designed with a variety of mechanisms to support attention, behavior arbitration, and motor expression (as discussed in section 7), these cognitive mechanisms are enhanced by emotion-inspired mechanisms that *further improve* Kismet's communicative effectiveness, its ability to focus its attention on relevant stimuli despite distractions, and its ability to prioritize goals to promote flexible behavior that is suitably opportunistic when it can afford to be yet persistent when it needs to be.

What about the external expression of emotion? Even if one were to accept the internal regulatory and biasing benefits of emotion-inspired mechanisms, do these need to be accompanied by social-emotive expression? Granted, it is certainly possible to use other information-based displays to reveal the internal state of robots: flashing lights, laser pointers, graphics, etc. However, people would have to learn how to decipher such displays to understand what they mean. Furthermore, information-based displays fail to leverage from the socio-affective impact and intuitive meaning that biological signals have for people.

Kismet's emotion system implements the style and personality of the robot, encoding and conveying its attitudes and behavioral inclinations toward the events it encounters. People constantly observe Kismet's behavior and its manner of expression to infer its internal state as they interact with it. They use these expressive cues as feedback to infer whether the robot understood them, its attitude about the interaction, whether they are engaging the robot appropriately, whether the robot is responding appropriately to them, etc. This helps the person form a useful mental model for the robot, making the robot's behavior more understandable and predictable. As a result, the person can respond appropriately to suit the robot's needs, and shape the interaction as he or she desires. It also makes the interaction more intuitive, natural and enjoyable for the person, and sustains his or her interest in the encounter.

In sum, although one could always argue that a robot does not *need* emotion-based mechanisms to address these issues, our point is that such mechanisms can be used to address these issues notably *better* than with cognitive-mechanisms alone. Furthermore, robots today are far from behaving and learning as intelligently, flexibly, and robustly as people and animals do with emotions. In our own work, we have shown that insights from emotion theory can be used to improve the performance of an autonomous robot that must pursue and achieve multiple goals in uncertain environments with people. With both cognitive and emotive systems working in concert, Kismet functions more adeptly—both from a decision-making and task-achieving perspective as well as from a social interaction and communication perspective.

As robot builders, we shall continue to design integrated systems for robots with internal mechanisms that complement and modulate its cognitive capabilities to improve the robot's overall performance. Several of these mechanisms may be close analogs to those regulatory, signaling, biasing, and other useful attention, value assessment, and prioritization mechanisms associated with emotion systems in living creatures. As a consequence, we will effectively be giving robots a system that serves the same useful functions that emotions serve in us—no matter what we call it. Kismet is an early exploration of these ideas and a promising first step. Much more work has yet to be done to more deeply explore, demonstrate, and understand the benefit of emotion-inspired mechanisms on intelligent decision-making, reasoning, memory, and learning strategies of autonomous robots. Improvement of these abilities will be critical for autonomous robots that will one day play a rich and rewarding part of our daily lives.

Acknowledgments

The principal author gratefully acknowledges the MIT Media Lab corporate sponsors of the Things That Think and Digital Life consortia for supporting her work and that of her students. Kismet was developed at MIT Artificial Intelligence Lab while working in the Humanoids Robotics Group of the second author. The development of Kismet was funded by NTT and DARPA contract DABT 63-99-1-0012.

References

Ackerman, B., Abe, J. & Izard, C. (1998), Differential emotions theory and emotional development: Mindful of modularity, *in* M. Mascolo & S. Griffen, eds, 'What develops in emotional development?', Plenum Press, New York, pp. 85–106.

Baron-Cohen, S. (1995), *Mindblindness*, MIT Press, Cambridge, MA.

Blumberg, B., Todd, P. & Maes, M. (1996), No bad dogs: Ethological lessons for learning, *in* 'Proceedings of the Fourth International Conference on Simulation of Adaptive Behavior (SAB96)', MIT Press, Cambridge, MA, pp. 295–304.

Breazeal, C. (2002*a*), *Designing Sociable Robots*, MIT Press, Cambridge, MA.

Breazeal, C. (2002*b*), 'Regulation and entrainment for human-robot interaction', *International Journal of Experimental Robotics* **21**(10-11), 883–902.

Breazeal, C. (2003*a*), 'Emotion and Sociable Humanoid Robots', *International Journal of Human-Computer Studies* **59**, 119–155.

Breazeal, C. (2003*b*), 'Emotive Qualities in Lip Synchronized Robot Speech', *Advanced Robotics* **17**(2), 97–113.

Breazeal, C. (2003*c*), 'Function Meets Style: Insights from Emotion Theory Applied to HRI', *IEEE Transactions on Systems, Man and Cybernetics: Part C*.

Breazeal, C. (2003*d*), 'Social Interactions in HRI: The Robot View', *IEEE Transactions on Systems, Man and Cybernetics: Part C.* forthcoming.

Breazeal, C. & Aryananda, L. (2002), 'Recognition of affective communicative intent in robot-directed speech', *Autonomous Robots* **12**(1), 83–104.

Breazeal, C. & Scassellati, B. (2000), 'Infant-like social interactions between a robot and a human caretaker', *Adaptive Behavior* **8**(1), 47–72.

Breazeal, C., Fitzpatrick, P. & Scassellati, B. (2001), 'Active vision systems for sociable robots', *IEEE Transations on Systems, Man, and Cybernetics: Part A* **31**(5), 443–453. K. Dautenhahn (ed.).

Brooks, R. (1986), 'A robust layered control system for a mobile robot', *IEEE Journal of Robotics and Automation* **RA-2**, 253–262.

Damasio, A. (1994), *Descartes Error: Emotion, Reason, and the Human Brain*, G.P. Putnam's Sons, New York, NY.

Darwin, C. (1872), *The Expression of the Emotions in Man and Animals*, John Murray, London, UK.

Dennett, D. (1987), *The Intentional Stance*, MIT Press, Cambridge, MA.

Ekman, P. (1992), 'Are there basic emotions?', *Psychological Review* **99**(3), 550–553.

Estrada, C., Isen, A. & Young, M. (1994), 'Positive affect influences creative problem solving and reported source of practice satisfaction in physicians', *Motivation and Emotion* **18**, 285–299.

Fernald, A. (1989), 'Intonation and communicative intent in mother's speech to infants: Is the melody the message?', *Child Development* **60**, 1497–1510.

Gould, J. (1982), *Ethology*, Norton.

Grosz, B. (1996), 'Collaborative Systems: AAAI-94 Presidential Address', *AI Magazine* **17**(2), 67–85.

Isen, A. (1999), Positive affect and creativity, *in* S. Russ, ed., 'Affect, creative experience, and psychological adjustment', Brunner-Mazel, Philadelphia, pp. 3–17.

Isen, A. (2000), Positive affect and decision making, *in* M. Lewis & J. Haviland-Jones, eds, 'Handbook of emotions, second edition', The Guilford Press, New York, chapter 27, pp. 417–435.

Isen, A., Rosenzweig, A. & Young, M. (1991), 'The influence of positive affect on clinical problem solving', *Medical Decision Making* **11**(3), 221–227.

Isen, A., Shalker, T., Clark, M. & Karp, L. (1978), 'Affect, accessibility of material and behavior: A cognitive loop?', *Journal of personality and social psychology* **36**, 1–12.

Izard, C. (1993), 'Four systems for emotion activation: Cognitive and noncognitive processes', *Psychological Review* **100**, 68–90.

Izard, C. (1997), Emotions and facial expressions: A perspective from differential emotions theory, *in* J. Russell & J. Fernandez-Dols, eds, 'The psychology of facial expression', Cambridge university press, Cambridge, pp. 57–77.

Izard, C. & Ackerman, B. (2000), Motivational, organizational and regulatory functions of discrete emotions, *in* M. Lewis & J. Haviland-Jones, eds, 'Handbook of Emotions', second edition edn, The Guilford Press, New York, chapter 16, pp. 253–264.

Lorenz, K. (1950), Part and parcel in animal and human societies, *in* K. Lorenz, ed., 'Studies in animal and human behavior', Vol. 2, Methuen, London, pp. 115–195.

Lorenz, K. (1973), *Foundations of Ethology*, Springer-Verlag, New York, NY.

Maes, P. (1991), Learning behavior networks from experience, *in* 'Proceedings of the First European Conference on Artificial Life (ECAL90)', MIT Press, Paris, France.

Maturana, H. & Varela, F. (1980), *Autopoiesis and cognition: the realization of the living*, D. Reidel, Boston.

McCulloch, W. & Pitts, W. (1943), 'A logical calculus of the ideas immanent in nervous activity', *Bulletin of mathematical biophysics* **5**, 115–133.

McFarland, D. & Bosser, T. (1993), *Intelligent Behavior in Animals and Robots*, MIT Press, Cambridge, MA.

Meltzoff, A. & Moore, M. K. (1997), 'Explaining Facial Imitation: A theoretical model', *Early Development and Parenting* **6**, 179–192.

Minsky, M. (1988), *The Society of Mind*, Simon and Schuster, New York.

Moore, B., Underwood, B. & Rosenhan, D. (1984), Emotion, self, and others, *in* C. Izard, J. Kagen & R. Zajonc, eds, 'Emotions, Cognition, and Behavior', Cambridge University Press, New York, pp. 464–483.

Norman, D. (2001), How Might Humans Interact with Robots, Keynote address to the DARPA/NSF Workshop on Human-Robot Interaction, San Luis Obispo, CA.

Picard, R. (1997), *Affective Computation*, MIT Press, Cambridge, MA.

Plutchik, R. (1991), *The Emotions*, University Press of America, Lanham, MD.

Premack, D. & Premack, A. (1995), Origins of human social competence, *in* M. Gazzaniga, ed., 'The Cognitive Neurosciences', Bradford, New York, NY, pp. 205–218.

Reeves, B. & Nass, C. (1996), *The Media Equation*, CSLI Publications, Stanford, CA.

Smith, C. & Scott, H. (1997), A componential approach to the meaning of facial expressions, *in* J. Russell & J. Fernandez-Dols, eds, 'The Psychology of Facial Expression', Cambridge University Press, Cambridge, UK, pp. 229–254.

Termine, N. & Izard, C. (1988), 'Infants' responses to their mothers' expressions of joy and sadness', *Develomental Psychology* **24**, 223–229.

Tinbergen, N. (1951), *The Study of Instinct*, Oxford University Press, New York, NY.

Tomkins, S. (1963), *Affect, imagery, consciousness: The negative affects*, Vol. 2, Springer, New York.

Trevarthen, C. (1979), Communication and cooperation in early infancy: A description of primary intersubjectivity, *in* M. Bullowa, ed., 'Before Speech: The Beginning of Interpersonal Communication', Cambridge University Press, Cambridge, UK, pp. 321–348.

Tronick, E., Als, H. & Adamson, L. (1979), Structure of early face-to-face communicative interactions, *in* M. Bullowa, ed., 'Before Speech: The Beginning of Interpersonal Communication', Cambridge University Press, Cambridge, UK, pp. 349–370.