SCHMOOZING WITH ROBOTS:

EXPLORING THE BOUNDARY OF THE ORIGINAL WIRELESS NETWORK

Cynthia Breazeal

MIT Artificial Intelligence Laboratory, 545 Technology SQ NE 43 – 938, Cambridge MA 02139, cynthia@ai.mit.edu

Anne Foerst

MIT Artificial Intelligence Laboratory, 545 Technology SQ NE 43 - 934, Cambridge MA 02139, annef@ai.mit.edu

INTRODUCTION

As robots take on an increasingly ubiquitous role in society, they must be easy for the average citizen to use and interact with. They must also appeal to persons of different age, gender, income, education, and so forth. This raises the important question of how to properly interface untrained humans with these sophisticated technologies in a manner that is intuitive, efficient, and enjoyable to use. Ideally, robots (and other interactive technologies) could participate in natural, human-style social exchange with their users.

How might we extend the boundary of humanity's social communication network (the "original wireless network") to include robots and other interactive technologies? To explore these issues, we present ongoing work to develop a socially interactive robot, which may someday serve as a physical avatar to improve and expand both human-machine interaction as well as some forms of inter-human relationships.

This seems to be quite a grandiose vision. It turns out, however, that the assumptions underlying this vision can be supported by scientific findings.

Humans treat computers like humans

A socially acting robot cannot be a successful intermediator between humans, unless the people in question accept the robot as competent. Much research has been carried out to investigate the human potential for anthropomorphizing gadgets, toys and computers. Can we assume this tendency to anthropomorphize interpersonal relationships? Within serious settings, it seems much more likely that people might reject a robotic intermediator.

However, recent research has shown that humans generally treat computers as they might treat other people, and it does not matter whether the people are computer experts, laypeople, or computer critics (Reeves/Nass 1996).[1] They treat computers with politeness usually reserved for humans, they are careful to not hurt the computer's 'feelings' by criticizing it; they feel good if the computer compliments them. In team play they are even are willing to side with a computer against another human if the human belongs to a different team. If asked before the respective experiment if they could imagine treating a computer like a person, they strongly deny it. Even after the experiment, they insist that they treated the computer as a machine and don't realize that they treated it as peer.

The main thesis is that the "human brain evolved in a world in which *only* humans exhibited rich social behaviors, and a world in which *all* perceived objects were real physical objects. Anything that *seemed*

to be a real person or place *was* real."(Reeves/Nass 1996, p.12). Our brains haven't evolved much further, they are still 'old' brains, and yet have to deal with twentieth-century technology. From these findings, we take as a working assumption that technological attempts to foster human-computer relationships will be accepted by a majority of people *if* the technological gadget displays rich social behavior. Evolution has hardwired our brains with an innate drive to interact in a social manner with others.

It is also important to note that human-style social interaction is different from that of ants, dogs, or other social animals. First, humans expect to share control with those whom they socially interact. This is a fundamental difference between interacting with others in the social world verses interacting with objects in the physical world. People rely on a variety of social mechanisms to share control with each other, such as turn taking and shared attention. As a consequence, social exchange between people is mutually regulated - as the interaction unfolds each participant's behavior responds and adapts to that of the other.

This dynamic is enriched by the manner in which humans can predict and socially influence the behavior of others through communicative acts. Much of this predictive power relies on each party being cooperative, open to communication, and subject to social norms. It also relies on each person viewing the other as an intentional being whose behavior can be explained and understood in terms of intents, beliefs, desires, emotions, and other mental states. Despite the human capacity to anthropomorphize non-human entities, human adults typically do not attribute goals and intents to inanimate physical objects in order to understand their behavior.

Finally, it is critical for each participant in the exchange to treat the other as a conspecific – to view the other as being "like me". Given such, the ability of each person to relate to and to empathize with the other helps each to predict and explain the other participant's behavior and to formulate appropriate responses based on this understanding. It also enables each to infer the intent behind the act of the other (e.g., what was said vs. what was meant). In short, adult human-style communication entails that each participant has a *theory of mind* of the other.

Infants learn sophisticated behavior through interaction with their caregivers

As robot designers, we are a long ways off from designing machines that can engage people in adult human-style communication, particularly within unconstrained social scenarios. However, by looking to the social development of human infants, it is possible to gain valuable insights into how these communicative competencies might be acquired. Our approach, therefore, entails constructing a robot with infant-like abilities and learning algorithms, and then have it acquire social competence through social experience, in rough analogy to what is understood about how children develop their own sociability.

There is another reason for focusing on newborns; most of what a human infant learns is acquired within an ongoing, dynamic, and social interaction process. This process begins immediately after birth with her parents, whom she depends upon for her survival. Hence the social experience to which all infants are naturally exposed is one in which one member of the interaction pair, the caregiver, is highly sophisticated and culturally competent, whereas the other, the infant, is culturally naive.

Soon after birth, babies respond to their caregivers in a well-coordinated manner. They seem to be born with a set of "pre-programmed" proto-social responses, which are specific to human infants. Their adaptive advantage seems to be their power to attract the attention of adults and to engage them in social interaction, the richness of which appears to be unique to the human species. For instance, newborns respond differently to people than other objects. Around people, they are more likely to vocalize and display facial expressions. They also show a preference for face-like stimuli over other sorts of pleasing stimuli, such as a bright red ball. Bateson (1979) argues that the infant's inability to distinguish separate words in her parent's vocalizations may allow her to treat their clauses as unitary utterances analogous to her own coos and murmurs. This allows the infant to participate in "dialogues" with them.

For the caregivers, their ability to present an appropriately complex view of the world to their infant strongly depends on how good they are at reading their infant's expressive and behavioral cues. It is

interesting how adults naturally engage infants in appropriate interactions without realizing it, and mothers seem to be instinctually biased to do so. For instance, *motherese* is an example of how adults simplify and exaggerate important aspects of language when speaking to infants (Bateson, 1979). By doing so, adults simplify the auditory processing task of the infant, making it easier for her to extract the meaningfully salient features of the adult's vocalizations. The use of exaggerated facial expressions is another example, where parents show extreme happiness or surprise during face-to-face exchanges with infants. In addition, there are a variety of biologically primed action patterns exhibited by the caregiver, which are matched to those of the infant (such as jostling the infants during pauses in feeding). These serve to encourage social interaction between mother and her infant, which fosters the infant's survival and continued development.

Given that the caregiver and infant engage in social interactions, there are a number of ways in which an infant limits the complexity of her interactions with the world. This is a critical skill for social learning because it allows the infant to keep herself from being overwhelmed or under stimulated for prolonged periods of time. For instance, the infant's physically immature state serves to limit her perceptual and motor abilities, which simplifies her interaction with the world. In addition, the infant is born with a number of innate behavioral responses that constrain the sorts of stimulation that can impinge upon her. Various reflexes such as quickly withdrawing her hand from a painful stimulus, evoking the looming reflex in response to an quickly approaching object, closing her eyelids in response to a bright light -- these all serve to protect the infant from stimuli that are potentially dangerous or too intense. In addition, whenever the infant is in a situation where her environment contains too much commotion and confusing stimuli, she either cries or acts to shut out the disturbing stimulation.

Humans are hardwired to react to facial stimuli

The caregiver is hardwired to react to the infant's signals; thus the infant can motivate and manipulate her caregivers to treat her properly and with care. Evidence for this hardwired response to a baby's signs can be found cross-culturally at the behavior level. Behavioral Scientists such as Eibl-Eibelsfeld (1970) argue that a cross-cultural '*Kindchenschema*' (baby-scheme) is embedded in the human brain; which biases humans to behave in a tender and caring way toward almost anything resembling an infant. Figure 1 shows examples of how this baby-scheme can be applied to dolls, animals, and cartoons.

People tend to react emotionally to someone or something 'cute' in this way; a natural behavior that is exploited by doll- and toy-producers as well as Hollywood designers… For our purposes as robot designers, it seems reasonable construct a robot with an infant-like appearance, which could encourage people switch on their baby-scheme and treat it as a cute creature in need of protection and care.

Figure 1: Examples of baby-scheme from [Eibl-Eibelsfeld, p. 33].

When designing our robot, we have made the engineering decision to focus on the face and head, ignoring other aspects of the body for now (arms, hands, legs, etc.). This decision was made due to the prominence of the face in social exchange. Indeed, our human reaction to faces is also hardwired into our brains (Cole 1998). Many experiments have shown that facial expressions are not only crucial for interaction, but that our brain has innate neural mechanisms which recognize faces and is preferentially activated if a face-like object is in view (shown in figure 2).

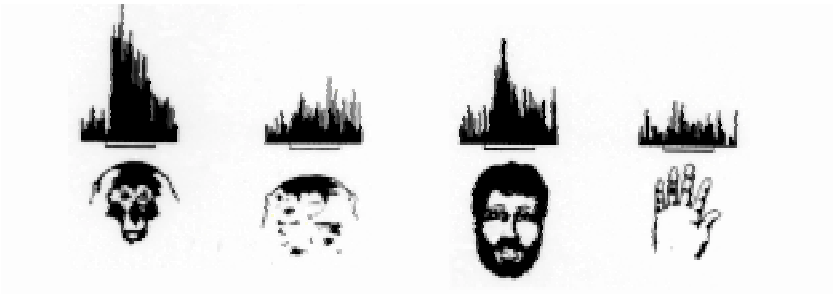

Figure 2: Preferential neural firing patterns to the presentation of a face in the inferior temporal cortex of the human brain (Churchland/Sejnowski 1992, p. 180).

In most of our social interactions, people depend on facial expressions for understanding the other and for being understood. If people lack the ability to display facial expressions (for instance with Möbius disease or after a stroke) they are not treated as equal partners within a human community. People interacting with them usually assume that these disabled people are numb, dumb, dull and demented. The primary facial expressions we use seem to be universal and cross-cultural. Despite obvious cultural differences while reacting in specific cultural settings and situations, the *primary* expressions such as anger, disgust, fear, sorrow, joy, etc. are shared across cultures. Numerous developmental psychologists have reported that infants both display them and respond to them.

Social interactions between caregivers and infants

Facial imitation takes place basically from the first day after birth (Meltzoff/Moore 1977); babies' "faces are the first part of the body that they take an interest in, not just to suck, but to express" (Cole, p.110). Trevarthen (1979) discusses how the wide variety of facial expressions displayed by infants are interpreted by the parent as indications of the infant's motivational state, are viewed as responses to their efforts to engage that infant, and encourage them to treat her as an intentional being. These expressive responses provide the parents with feedback, which they use to carry the dialog along.

Tronick and colleagues (1979) identify five phases that characterize social exchanges between three-month-old infants and their caregivers: *initiation, mutual-orientation, greeting, play-dialog* and *disengagement*. Each phase represents a collection of behaviors that mark the state of the communication. Not every phase is present in every interaction. For example, a greeting does not ensue if mutual orientation is not established. Furthermore, a sequence of phases may appear multiple times within a given exchange, such as repeated greetings before the play-dialog phase begins. To the parents, this interaction feels like a dialogue (sometimes called a proto-dialog) carried out with their child even though it is non-language based.

The early proto-social responses exhibited by infants are a close enough approximation to the adult forms that caregivers immediately interpret their babies' reactions by a process of adultomorphism. Simply stated, parents assume their infant to be fully socially responsive; with wishes, intentions, and feelings which can be communicated to others and which must be respected within certain limits. Events that may at first be the result of automatic action patterns, or may even be spontaneous or accidental, are endowed with social significance by the caregivers. At such an early age, Kaye (1979) and Newson (1979) point out that it is the parent who supplies the meaning to the exchange, and it is the mechanism of flexible turn taking that allows them to maintain the illusion that a meaningful exchange is taking place. By assuming that their infant is attempting some form of meaningful dialog, and by crediting her with having thoughts, feelings, and intentions like all other members of society, they impute meaning to the exchange in a consistent and reliable manner. By doing so, they establish a "dialog" with their infant, from which the communication of shared meanings gradually begins to take place. It is the consistency of these exchanges that allows her to learn the meaning her acts have for others.

Summary

To summarize, humans are tuned for social interaction and gadgets that behave socially are treated like people (in a restricted sense). Social interaction is especially important between parents with their offspring; babies cannot acquire sophisticated social skills and understanding without it. At birth, they are bootstrapped into social exchanges with their parents by activating innate proto-social responses. This motivates the caregiver to treat the infant as if she was already a social being, complete with her own beliefs, intents, and desires. From these exchanges, the infant develops her sociability.

Expressive display, plays an important part in most of these early interactions. Even in adulthood, facial displays are very important in interpersonal communication. Humans with non-moveable faces are not treated as "full-fledged" persons (even though a computer without any face is treated as such; which can be explained with different expectations). Basic facial expressions seem to be cross-culturally consistent and humans are extremely sophisticated in reading faces.

TOWARD HUMAN-STYLE INTERACTION BETWEEN HUMANS AND ROBOTS

We want to build a robot that will be able to interact with people in a human-like way. From the previous section, we have a set of important design constraints:

- *Issue I*: the robot should have a cute face to trigger the 'baby-scheme' and motivate people to interact with it, to treat it like an infant, and to modify their own behavior to play the role of the caregiver (e.g. using motherese, exaggerated expressions and gestures).

- *Issue II*: The robot's face needs several degrees of freedom to have a variety of different expressions, which must be understood by most people. Its sensing modalities should allow a person to interact with it using natural communication channels.

- *Issue III*: The robot should be pre-programmed with the basic behavioral and proto-social responses of infants. This includes giving the robot the ability to dynamically engage a human in social. Specifically, the robot must be able to engage a human in proto-dialogue exchanges.

- *Issue IV*: The robot must convey intentionality to bootstrap meaningful social exchanges with the human. If the human can perceive the robot as a being "like-me", the human can apply her social understanding of others to predict and explain the robot's behavior. This imposes social constraints upon the caregiver, which encourages her to respond to the robot in a consistent manner. The consistency of these exchanges allows the human to learn how to better predict and influence the robot's behavior, and it allows the robot to learn how to better predict and influence the human's behavior.

- *Issue V*: The robot needs regulatory responses so that it can avoid interactions that are either too intense or not intense enough. The robot should be able to work with the human to mutually regulate the intensity of interaction so that it is appropriate for the robot at all times.

- *Issue VI*: The robot must be programmed with a set of learning mechanisms that allow it to acquire more sophisticated social skills as it interacts with its caregiver.

Kismet, Our Interactive Infant Robot

Figure 3: Appearance-wise, Kismet has been designed to have an infan appearance according to the 'baby-scheme' of Eibl-Eibelsfeld (address design Issue I). Specifically, Kismet has large eyes, a demure chin, lips suggest the ability to suck, and the suggestion of a high forehead.

Kismet is a robotic head consisting of a variety of perceptual and motor systems specialized for human-style communication (see figure 3). The robot is designed according to the 'baby-scheme' of Eibl-Eibelsfeld to encourage people to treat it like an infant and to play typical infant-caregiver games with it. For the caregiver, this might include using excessive prosody when speaking to the robot (a.k.a. "*motherese*"), exaggerating facial expressions and gestures, and playing simple imitation-based games such as mimicking facial expressions and babbling. These infant-adapted interactions are of great benefit to the robot just as they are to the infant --- acting to exaggerate the most salient aspects of the interaction to make them easier for the robot to perceive and interpret.

When face-to-face, people use a wide variety of sensory and motor modalities to communicate. To date, research efforts have focused primarily on the perception of human gesture and speech to convey task-based information to interactive technologies. During social exchange, however, being aware of the other's *motivational state* is also critical. Humans use numerous affective cues (e.g., facial expressions, vocal prosody, body posture) as well as social cues (e.g., direction of gaze, feedback gestures such as nods of the head, raising eyebrows) to infer the intents, beliefs, and wishes of the other. To enable Kismet to perceive these cues (addressing *Issue 2*), the robot is equipped with a stereo active vision system, and an active stereo auditory system. Each eye has a color CCD camera situated behind the pupil with a 5.6*mm* focal length lens. There is also a color CCD camera mounted on the side of the head with a 2.0*mm* focal length lens, providing the robot with a wide peripheral field of view. Each ear has a small microphone, both mounted on a movable head. The robot has a proprioceptive sense via motor encoders that allow it to sense its own motion.

The robot's output modalities enable the robot to deliver a variety of important social cues. To be able to direct its gaze, Kismet's eyes have three degrees of freedom (DoF): each eye can pan independently and both are coupled to a common tilt DoF. These axes enable Kismet to direct its gaze to salient environmental stimuli such as the caregiver's face. Kismet's neck also has three degrees of freedom: a pan DoF, an upper tilt DoF and a lower tilt DoF. These DoFs allow the robot to orient its head, to perform communicative gestures (such as nodding and shaking of the head), and to posture itself expressively (approaching something of interest by leaning forwards, withdrawing from something undesirable by leaning backwards). A speech synthesizer enables the robot to produce vocal utterances, similar to a babbling infant, and to adjust the prosodic contour of its utterance.

Kismet's Expressive Face

Kismet's head is also embellished with facial features for emotive expression (addressing *Issue II*). Currently these facial features include controllable eyebrows, ears, eyeballs, and eyelids, and a mouth with an upper lip and a lower lip. The robot is able to show a wide variety of recognizable facial expressions (as shown in figure 5), as well as control the facial features for communicative displays. Instead of being programmed with a discrete set of facial expressions, the algorithm allows the robot to display a continuous range of expressions of varying intensity (using a scheme similar to Russell's (1980) for characterizing emotions in humans). Briefly, each of Kismet's expressions is represented as a point within an *expression-space.* The space consists of three dimensions: *arousal* (i.e. high, neutral, or low), *valence* (i.e. positive, neutral, or negative), and *stance* (i.e. open/accepting, neutral, or closed/rejecting). Each dimension has a characteristic facial posture for each extreme (see figure 4), and these six facial postures span the space of all possible expressions that Kismet can generate.

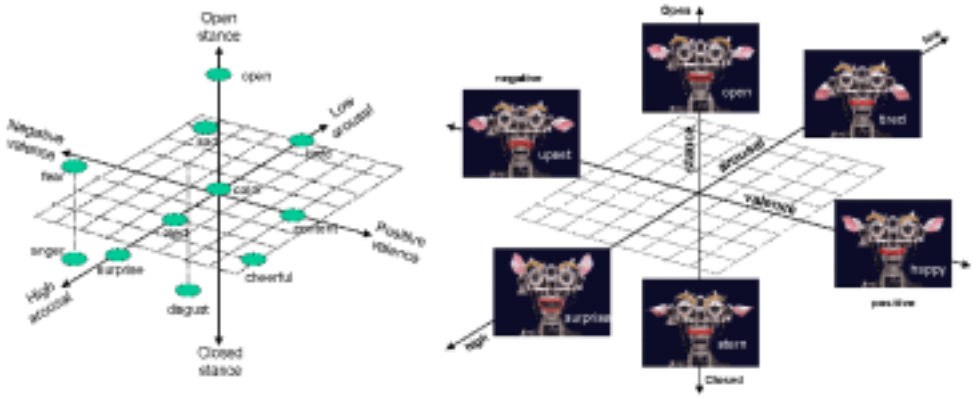

Figure 4: The figure to the left illustrates how various emotional states can be characterized in terms of arousal, valence, and stance parameters. The affect-space scheme we use is a variation on that of Russell, whose space consists of arousal, valence, and potency dimensions. The figure to the right illustrates the basis set of postures of the robot's expression-space. An expression for each point in affect-space is generated by algorithmically combining these basis postures.

A web-based experiment was conducted to determine if non-technical people, unfamiliar with the robot, would be able to correctly recognize its facial expressions [*http://www.spiritualityhealth.com/news/robots.html*]. Fifteen subjects, 86% of which are Caucasian middle-aged women of non-technical background, were simply asked to label the expressions shown in figure 5 (the web site images are in color). The subjects were told that the top figure shows the robot displaying a calm/neutral expression. The subjects were then asked to label the remaining images. The most frequently used label is shown below each image. The most representative label of the average response is shown to the right if different from the most commonly used label. To compute the most representative label for each image, all responses for that image were systematically decomposed into their arousal, valence, and stance constituents. A value of *+10* was assigned to each constituent characterized as high arousal, positive valence, or open stance. A value of *–10* was assigned to each constituent characterized as low arousal, negative valence, or closed stance. A zero value was given to constituents characterized as having neutral arousal, neutral valence, or neutral stance. We computed a weighted average of these arousal, valence, and stance, which is shown below each image as the trio: (arousal, valence, stance). Hence, the most representative label is the response that most closely corresponds to this arousal, valence, and stance trio.

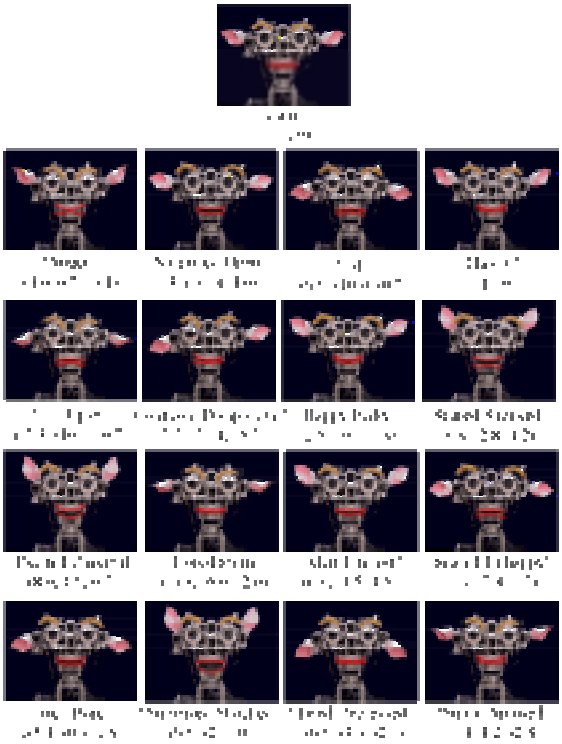

Figure 5: With the expression-space scheme, Kismet can display a wide range recognizable expressions. This flexibility allows Kismet to display a distinct and ea readable expression consistent with its affective state.

The images to the right were presented t the fifteen subjects. The scores are show under each image (see text). The subject were easily able to infer arousal states (excited vs. subdued) and valence states (good/happy vs. bad/upset) from the images. Stance was more difficult for the to infer, yet seemed to be the distinguish factor of similar labels for different imag For instance, note the "scared/shocked" the "scared/unhappy" distinction in the column of images.

Behavior Control Architecture

The organization and operation of Kismet's control software is heavily influenced by concepts from psychology, ethology, and developmental psychology, as well as the applications of these fields to robotics (Brooks et al. 1998). The overall system is implemented as an agent-based architecture similar to those presented in (Blumberg 1996) and (Minsky 1988). The system architecture consists of five subsystems: the *perception system*, the *motivation system*, the *attention system*, the *behavior system*, and the *motor system*. The perception system extracts low-level perceptual features from the world as filtered through the robot's sensors. The motivation system maintains the robot's internal state in the form of computational models of drives and emotions. The attention system determines the saliency of stimuli based upon perceptual and motivational influences. The behavior system implements various behavioral and proto-social responses, and is responsible for activating them using a behavior arbitration scheme as conceptualized in (Tinbergen 1951) and (Lorenz 1973). Finally, the motor system realizes these behaviors as facial expressions and other motor skills. The technical details of these systems have been reported in a number of publications (Breazeal 1998, Breazeal&Scassellati 1999, Breazeal 1999). For the purposes of this paper, we simply highlight the requisite social skills (*Issues I* --- *V* as outlined above) that Kismet performs as generated by its control system.

Proto-Social Responses

Kismet's control software is designed so that Kismet exhibits those infant-like responses that most strongly encourage people to interact with it as if it were an infant and to attribute intentionality to it (addressing *Issues III* and *IV*). To convince a human that Kismet has internal goals, beliefs, and desires, the robot's subjective internal states must express themselves (as facial expressions or body postures) in response to external events, just as for human infants. Acts that make subjective processes overt include focusing attention on objects, orienting to external events, handling or exploring objects with interest, and so forth. These responses can be divided into four categories. *Affective responses* allow the human to attribute feelings to the robot. *Exploratory responses* allow the human to attribute curiosity, interest, and desires to the robot, and can be used to direct the interaction to objects and events in the world. *Protective responses* keep the robot away from damaging stimuli and elicit concerned and caring

responses from the human. *Regulatory responses* maintain a suitable environment that is neither too overwhelming nor under-stimulating.

The robot's internal state (emotions, drives, concurrently active behaviors, and the persistence of a behavior) combines with the perceived environment (as interpreted thorough the perception and attention systems) to determine which behaviors become active. Once active, a behavior can influence both what the robot does (by influencing motor acts such as gaze direction and head orientation) and how that action is expressed through current facial expression (by influencing the arousal, valence, and stance aspects of the emotion system). Many of Kismet's behaviors are motivated by emotions as proposed by Plutchik (1984), who summarizes from an evolutionary perspective under what conditions certain primitive emotions and behavioral responses are aroused in animals as well as humans. These emotive responses have been adapted for Kismet as shown in table 1. In this way, Kismet's emotional responses mirror those of biological systems and thereby seem plausible to a human observer (addressing *Issue IV*). Like the human infant, it is not necessary that our robot initially understand the significance of its actions. The consistent interpretation given by the caregiver will allow the robot to eventually learn of the significance of its actions.

Prototype	Function of the Behavior	Emotion Associated	Activation Conditions for Kismet
Incorporation	Accept environmental stimulus	acceptance	Acceptance of a desired stimulus
Rejection	Get rid of something harmful already accepted	disgust	Appearance of an *un*desired stimulus
Protection	Avoid being destroyed	fear	Appearance of a threatening, overwhelming stimulus
Deprivation	React against important loss	sorrow	Loss of a desired stimulus
Orientation	React to a new or strange object	interest	Appearance of new, *salient* stimulus
Exploration	Explore environment	boredom	Need of a desired yet absent stimulus
Reward	Reinforce beneficial behavior	joy	Success in achieving goal of active behavior
Destruction	Remove barrier to achieve some need	anger, frustration	Delay in achieving goal of active behavior

Table 1: Kismet's affective and behavioral responses to events in the world are adapted from the work of Plutchik. The most significant difference is the substitution of Kismet's "reward" prototype response for Plutchik's "reproduction" prototype response.

The four categories of behavior (affective, exploratory, protective, and regulatory) have been integrated into Kismet's behavior system in the form of eight prototype behavioral responses (incorporation, rejection, protection, deprivation, orientation, exploration, reward, and destruction). The robot displays affective responses by changing facial expressions in response to stimulus quality and its current internal state. A second class of affective response results when the robot expresses preference for one stimulus type. Exploratory responses include visual search for desired stimuli and maintenance of mutual regard. Kismet currently has a single protective response, which is to turn its head and look away from noxious or overwhelming stimuli. Finally, the robot has a variety of regulatory responses including: biasing the caregiver to provide the appropriate level of interaction through expressive feedback; the cyclic waxing and waning of affective, attentive, and behavioral states; habituation to unchanging stimuli; and generating behaviors in response to internal motivational requirements. These behavioral responses encourage the human to treat a robot as an intentional creature and to establish meaningful communication with it.

Dynamics of infant-like social interaction

As figure 6 shows, Kismet's control architecture produces interaction dynamics similar to the five phases of infant-caregiver social interactions as described by Tronick, Als, and Adamsen 1979) (addressing *Issue III*). The interaction begins with an initiation phase, followed by a mutual orientation phase, which

results in a greeting phase. A play-dialog phase ensues where the caregiver engages the robot with her face and then a toy. A disengagement phase ends the interaction run. These dynamic phases are not explicitly represented in the software architecture, but emerge from the interaction of the control system with the environment. By producing proto-social behaviors that convey intentionality, the caregiver's natural tendencies to treat the robot as a social agent cause her to respond in characteristic ways to the robot's overtures. This reliance on the external world produces dynamic behavior that is both flexible and robust. We refer the interested reader to (Breazeal&Scassellati 1999) for an in-depth presentation of the implementation details of these responses as well as the results of further experiments.

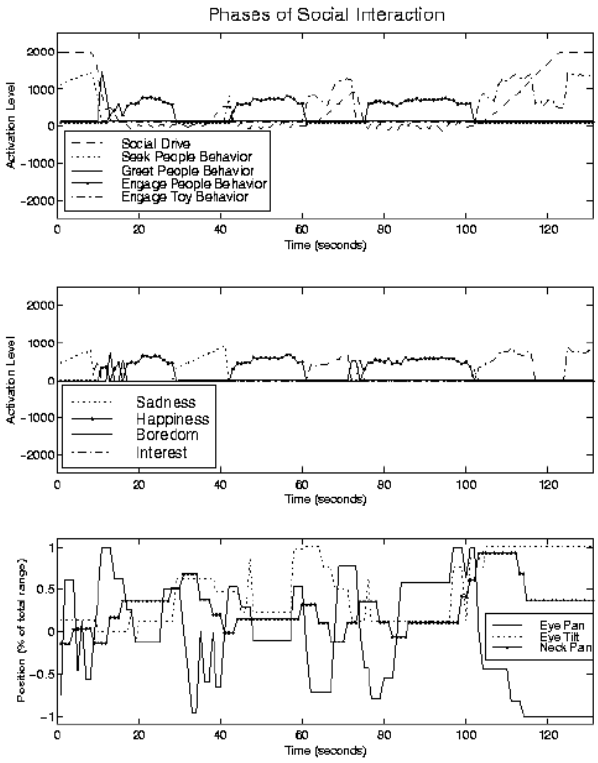

Figure 6: Kismet's dynamic responses during face-to-face interaction with a caregiver.

Kismet is initially looking for a person a displaying sadness (the initiation phase) The robot begins moving its eyes lookin for a face stimulus (t<8). When it finds t caregiver's face, it makes a large eye movement to enter into mutual regard (10). Once the face is foveated, the robo displays a greeting behavior by wiggling ears (t_ 11), and begins a play-dialog phase of interaction with the caregiver (t>12). Kismet continues to engage the caregiver until the caregiver moves outsi the field of view (t 28). Kismet quickly becomes sad, and begins to search for a face, which it re-acquires when the caret returns (t 42). A final disengagement phase occurs (t 100), when the robot attention shifts to a toy.

Mutually Regulating Social Interaction

Similar to a human infant, Kismet's emotional responses during interactive play provide important social cues that the caregiver uses to assess how to satiate the robot's drives (a.k.a. its "needs"), and how to carefully regulate the complexity of the interaction (addressing *Issue V*). The former is critical for the robot to learn how its actions affect its caregiver, and the later is critical for establishing and maintaining a suitable learning environment where Kismet is neither bored nor over-stimulated. Kismet has a number of drives: two of which are a *social drive* which represents the robot's need to be stimulated by people, and a *stimulation drive* which represents the robot's need to be played with using toys. In general, as long as the robot's drives remain satiated, the robot displays an interested or content expression. However, as a drive increases in intensity (due to a lack of the desired sort of stimulation) it becomes more urgent for the robot to satiate that particular need. Consequently, the robot appears increasingly depressed and is motivated to act in ways to acquire the desired stimulation. In contrast, if the robot is being over-stimulated by the interaction, then the robot appears increasingly distressed and is motivated to act in ways to reduce the intensity of the stimulation. These visual cues (both the observable behavior of the robot as well as its facial expressions) tell the human that all is not well with the robot, and whether the human should intensify the interaction, diminish it, or maintain it at its current level.

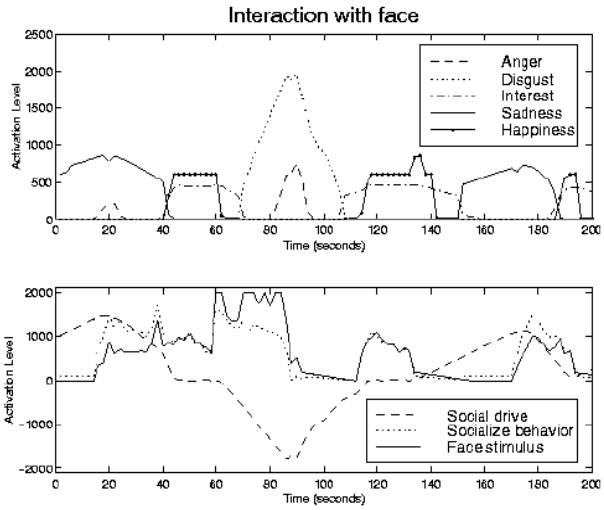

Figure 7: Due to a low intensity of human interaction from *0<=t<= 15*, the robot becomes increasingly "sad" as the social remains unsatiated (represented by large positive values). The robot's expression of sadness continues to increase, until the human finally responds by intensifying the interaction. This satiates the social drive (its magnitude tends towards zero), and the human sees the robot's "sadness" decay until it appears "interested" (from *45<=t<=60*). In contrast, from *60<=t<=90* the robot acquires more "asocial" tendencies when the interaction is too intense and the social drive moves toward the overwhelmed end of the spectrum (as represented by large negative values). As this drive leaves the homeostatic range, the robot becomes increasingly "disgusted" and its expression of "disgust" intensifies over time. When the social drive reaches a fairly large negative value of *-1500*, the robot also begins to display signs of "anger", and the human backs off the interaction. This causes the social drive to return to the homeostatic range and the robot re-establishes an "interested", "happy" appearance.

Figure 7 gives a brief illustration of the robot's behavior when interacting with a human. It illustrates the influence of the social drive on the robot's motivational and behavioral state when interacting with a human simply by face-to-face contact. It demonstrates how the robot's emotive cues are used to regulate the nature and intensity of the interaction, and how the nature of the interaction influences the robot's behavior. The result is an ongoing "dance" between robot and human aimed at maintaining the robot's drives within homeostatic bounds. If the robot and human are good partners, the robot remains "interested" and/or "happy" most of the time. Implementation details of this mechanism and more extensive experiments have been reported previously in (Breazeal 1998)].

Summary

To summarize, we have briefly presented ongoing work on building an infant-like robot that engages humans in social interactions. The physical appearance of the robot encourages people to treat the robot in an infant-like way. The robot's sensing and motor abilities are designed to mirror the natural communication channels of humans. Of particular importance are the robot's expressive abilities and its affective responses to the caregiver's behavior. These play a prominent role in emulating those social exchanges between caregivers and infants. When engaging a human, the dynamics of interaction between robot and human mirrors that of infant-caregiver pairs. We have also shown how the robot's expressive cues enable it to mutually regulate the social exchange so that the robot is neither overwhelmed nor under stimulated. This skill has important implications during socially situated learning episodes.

Addressing *Issue VI* (developing the socially grounded learning mechanisms) is the focus of ongoing work. In particular, the implementation of imitative learning abilities for facial imitation and vocal mimicry is

underway. Given the importance of imitation-based learning for human infants in acquiring social competence, implementing such learning abilities is of primary importance for future work with Kismet.

APPLICATIONS

The cuteness of Kismet triggers the 'baby-scheme' in humans, which suggests that many people will perceive Kismet-like technologies as toys. On the other hand, we know that Kismet can evoke strong emotional responses from people. The ambiguity in the perception of Kismet (ranging from "just a toy" to "disconcerting!") can be used in several interesting applications.

Physical telecommunication:

As researchers unfold the importance of non-verbal cues in social interactions, the importance of many subconscious body-routines and paralinguistic cues are becoming increasingly recognized.[2] Embodied reactions toward a speaker such as widening eyes if the statement was important for the listener will change the speaker's performance, and if perceived as positive, will often improve her thought processes and her creativity. Various gestures are used to regulate the rate of the conversation (e.g. feedback nods of the head are used to indicate that the listener is following the conversation). Other social cues (such as gaze direction, shared attention, and deictic pointing) are frequently used by people when communicating about objects in the physical world. However, in network interaction using flat displays, many of these embodied cues (particularly those expressed in 3D space) are lost, thus diminishing the richness of the communicative exchange. Kismet-like technologies could be used as a physical avatars in network communities to provide the embodied, non-verbal cues of the respective speaker (such as pointing to objects or shared attention), in the *same* physical space of the listener, thus supporting physical-based interactions between people in different physical spaces.

For this application, the physical avatar would not require elaborate social intelligence. Instead, the avatar could be largely tele-operated to convey non-verbal aspects of the communication to the other dialogue partner(s). Nonetheless, such an application would still require the designers of this technology to address challenging perceptual and body-mapping issues.

Tool for People with Disabilities:

Facial expressions seem particularly important for sending non-verbal information to dialogue partner(s). As mentioned earlier, people without expressive faces are taken less seriously and are approached in less meaningful ways by their relatives and friends. In cases of stroke patients who may have lost control over their facial muscles, or people with a nervous disorders which hinders them from creating facial expressions (eg. Möbius disease), Kismet-like technologies could serve as an embodied communication tool, serving as a physical avatar for people sharing the same physical space and time. If the avatar technology could sense the affective state of the person through other channels besides just facial movement, the avatar could display the intended facial display.

Fortunately, research is underway to measure the affective state of people. For instance, Roz Picard and her collaborators have developed several systems to measure people's emotional states (Picard 1997). Currently she is training platforms to recognize affective states using a variety of different sensors placed on different regions of the body (e.g. face, arm, finger, and waist). The sensors measure respiration level, pulse, and skin conductance, all of which change with changes in emotional states. Such technologies could be used to provide the emotional state of a patient as input to the embodied avatar. In contrast to the previous application, in this scenario the physical avatar would need some form of semi-autonomous social intelligence to display the appropriate non-verbal cue after interpreting the sensor data and understanding the social context (the robot might add smiles, feedback head-nods, etc. to the conversation).

Teaching Platform for Expectant Parents:

Kismet-like platforms could be used to teach expectant couples or mothers – especially teens. These robotic "dolls" could give people practice in reading and understanding their to-be-infant's attempts at communication, and subsequently teach the soon-to-be parents how to react sensibly. Such sensitivity to the infant is critical for forming proper attachment between parent and child, which has tremendous consequences to the infant's own development. The subject could learn certain "to do's" (such as interacting with the baby socially, engaging it in the exploration of the world, and treat it as a fully aware and vulnerable person). They could also learn certain "not to do's" (such as over-stimulating the infant, expecting too much, or being too fast for the baby). They will also experience the feeling of full-time responsibility and commitment.

For this application, the robotic "doll" should be able to perform infant-like social skills and behavioral responses (several of which have been implemented on Kismet). The robot would also require socially grounded learning abilities, such as imitation-based learning, and the ability to form an attachment to the caregiver. These adaptive abilities would allow the robot "doll" and caregiver to learn and grow together, providing rich interpersonal simulations in preparation for the infant's arrival.

Encouragement for Social Interaction:

There are cases where children are socially disabled. In some cases this may be a clinical condition (e.g. autism[3]), in other cases it might be a result of poor early childhood experiences (e.g. for children raised in understaffed orphanages with an insufficient caregiver-child ratio). Kismet-like technologies could help promote social, non-verbal interaction for these children. The robotic "pet" could either encourage children to interact with it socially, or it could encourage the child to interact socially with other people.

To do this properly, the robotic "pet" needs some rudimentary social skills to engage children in a social way. Both the child and the robot could develop a simple relationship over time as the child learns to reach out and enter relationships with other children. The robot would also need a fair amount of social sophistication and intelligence to motivate the child to interact with it over the long term. Because of the level of sophistication of social skills such a device would require, this application remains a long-range vision.

EPILOGUE

In conclusion, we would like to speculate about what a realized socially intelligent robot might mean for our society. Despite its cuteness, Kismet can create fear and can make people feel uncomfortable.[4] It seems to be disconcerting idea for them to interact with a socially intelligent machine that also triggers our baby-responses. Often, people search for qualitative differences between Kismet and themselves in order to preserve their sense of human specialness.

This negative reaction toward socially intelligent robots is justified insofar that indeed many AI researchers have claimed that a realized artificially intelligent device (that emulates human behavior and competencies) would prove that humans are nothing but machines. This statement, of course, is circular - it only makes sense to attempt to create an artificially intelligent machine if one assumes that intelligence can be completely analyzed and implemented in mechanistic. Such a machine, therefore, cannot prove this assumption. Unfortunately, however, many people assign scientific insights and technological gadgets enormous power over their own self-understanding – often eliciting frightened to Kismet and the notion of socially intelligent robots.

Our reaction to these findings is twofold. For one, Kismet fulfills the 'baby-scheme' and is not as threatening as other humanoid robots. But even more important is our attempt to present Kismet to a very broad audience to give people the sense that we do not impose this technology upon society. We present Kismet on TV (Nightline, Odyssey Channel, etc.), in the popular press (Discover, Forbes, etc.), and have run an experiment on the website of "Spirituality & Health" (a magazine that does not necessarily attract people who feel positive toward technology). We work to inform as many people as

possible in order to start a broader societal discussion about the acceptance of socially intelligent technologies and their possible design.

When we engage people in discussions about what these machines mean for us, where to use them, and of our own self-understanding, the question of Kismet's personality and personhood often comes up. Would we ever consider Kismet to be a person, a member of our community that cannot be switched off anymore? What would these conditions be?

This makes us think about the people in our own communities anew; we ask ourselves why we accept other people into our communities. Careful analysis of sociological data (e.g. Reeves&Nass) reveals that acceptance is rarely based on hard empirical facts. Instead, it is based on emotions, empathy, sympathy and the hardwired functions in the human brain. If a significant number of people were to accept a socially intelligent robot into their group, and treat the robot as and equal member of it, could that robot then become part of the human community? For instance, if someone within our own research group (someone who knows exactly how Kismet functions) at some point decides that Kismet has stepped over the 'magic' threshold that distinguishes an object from a person, would that be valid and convincing for other people as well?

Kismet is built on the principle that interacting with and being a part of the world is crucial for any existence of intelligence. As Kismet inspires social treatment of itself, we might rediscover the personhood of other humans. We might learn to appreciate, that ultimately it is we who will decide what the differences between our selves and our creations really are.

REFERENCES

Bateson, M. (1979). The epigenesis of conversational interaction. In M. Bullowa (Ed.), *Before Speech*, Cambridge University Press, 63-77.

Blumberg, B. (1996). Old tricks, new dogs: ethology and interactive creatures, PhD. Thesis, MIT.

Breazeal, C. (1998). A motivational system for human-robot interaction. In*: Proceedings of the fifteenth national conference on artificial intelligence*, Madison, WI, 54-61.

Breazeal, C. & Scassellati, B. (1999). A context-dependent attention system for a social robot. In *Proceedings of the 1999 international joint conference on artificial intelligence*, Stockholm, Sweden.

Breazeal, C. (1999). Robot in society: friend or appliance. In *Proceedings of the 1999 autonomous agents workshop on emotion-based architectures*, Seattle, WA, 18-26.

Breazeal, C. & Scassellati, B. (1999). How to build robots that make friends and influence people. *In Proceedings of the 1999 IEEE/RSJ international conference on intelligent robots and systems*, Kyonjiu, Korea.

Brooks, R., Breazeal, C., Irie, R., Kemp, C., Marjanovic, M., Scassellati, B., & Williamson, M. (1998). Alternative essences of intelligence. In Proceedings of the fifteenth national conference on artificial intelligence, Madison, WI, 961-967.

Churchland P. & Sejnowski, T. (1992*). The Computational Brain*. Cambridge, MA: MIT-Press.

Cole, J. (1998). About Face. Cambridge, MA: MIT-Press.

Eibl-Eibelsfeld, I. (1970). Liebe und Hass: Zur Naturgeschichte elementarer Verhaltensweisen. Munic, Germany: Piper.

Kaye, K. (1979). Thickening thin data: the maternal role in developing communication and language. In M. Bullowa (Ed.), *Before Speech*, Cambridge University Press, 191-206.

Lorenz, K. (1973). *Foundations of ethology*. Springer-Verlag.

Meltzoff, A. & Moore, M.K. (1977). Imitation of facial and manual gestures be human neonates. *Science,* 198, 75-78.

Minsky, M. (1988). *Socitey of Mind*. Simon and Schuster.

Newson, J. (1979). The growth of shared understandings between infant and caregiver. In M. Bullowa (Ed.), *Before Speech*, Cambridge University Press, 207-222.

Picard, R. (1997). *Affective Computing*. Cambridge, MA: MIT-Press.

Plutchik, R. (1984). Emotions: a general psycho-evolutionary theory. In K. Scherer & P. Elkman (Eds.),

Approaches to Emotion. New Jersey: Lawrence Erlbaum Associates, 197-219.
Trevarthen, C. (1979). Communication and cooperation in early infancy: a description of primary intersubjectivity. In M. Bullowa (Ed.), *Before Speech*. Cambridge University Press, 321-348.
Tronick, E., Als, H., & Adamson, L. (1979). Structure of early face-to-face communicative interactions. In M. Bullowa (Ed.), *Before Speech*. Cambridge University Press, 349-370.
Reeves, B. & Nass, C. (1996). *The Media Equation: How People Treat Computers, Television, and New Media Like Real People and Places*. Cambridge UK: Cambridge University Press.
Russell, J. (1980). A circumplex model of affect. *Journal of personality and social psychology*.
Tinbergen, N. (1951). *The study of instinct*. Oxford University Press.

---

[1] Interestingly enough, Reeves and Nass used particularly 'un-social' computers which were old, slow, with bad b/w monitors; they wanted to avoid the "Tamagotchie" effect by faking social responses in the computer. The people nonetheless couldn't help but treat the computers as persons.
[2] Justine Cassell (MIT Media-Lab) has done extensive research on non-verbal communication for her human-computer interfaces; most of our information comes from personal discussions with her and from reading her newest (and yet unpublished) work.
[3] See the recent work of Kerstin Dautenhahn, using robotic technologies to foster the social development of autistic children.
[4] Whenever Dr. Foerst presents Kismet either for the academic field of "religion & science" or in teaching ministers, some of the listeners will express their negative reaction in the Q+A period.