Robot in Society: Friend or Appliance?

Cynthia Breazeal *
Massachusetts Institute of Technology
Artificial Intelligence Laboratory
545 Technology Square, Room 938
Cambridge, MA 02139 USA
email: cynthia@ai.mit.edu

Abstract

This paper discusses the role that synthetic emotions could play in building autonomous robots which engage people in human-style social exchange. We present a control architecture which integrates synthetic emotions and highlight how they influence the internal dynamics of the robot's controller — biasing attention, motivation, behavior, learning, and the expression of motor acts. We present results illustrating how this control architecture, embodied within an expressive robot and situated in a social environment, enables the robot to socially influence its human caregiver into satisfying its goals.

1 Robot in Society

The number of service robots performing useful tasks such as delivering office mail or hospital meals is growing and may increase dramatically in the near future (Klingspor, Demiris & Kaiser 1997). In Japan, *MITI* and *AIST* have launched a large humanoid robot research initiative. Possible applications include bringing humanoid robots into the household to serve as nursemaids and butlers for Japan's growing elderly population (Inoue 1998). As the tasks autonomous robots are expected to perform become more complex, so do the robots themselves. 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.

This paper examines several key issues in building socially intelligent autonomous robots (SIARs) that engage humans in unconstrained social interactions. Going beyond the view of human-robot communication as transferring task-based information via intuitive communication channels, we argue that creating SIARs will require designers to address the emotional and inter-personal dimensions of social interaction. We illustrate one possible implementation of this methodology by describing ongoing work on Kismet, our socially situated robot. Affective influences play a prominent role in Kismet's control architecture, biasing the internal dynamics of several sub-systems including perception and attention, motivation, behavior, motor expression, and ultimately mechanisms for social forms of learning. We present results illustrating how this architecture enables Kismet to engage in believable social interaction with a person, and how it enables Kismet to influence its social world to maintain its internal agenda.

2 Robot Appliances

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. In designing effective human-robot interfaces, one must carefully consider the relevant human-factors to ensure a good fit. For instance, how are people naturally inclined to interact with and use these technologies, and how does the interaction impact the person using it?

Recognizing the importance of the human-robot interface issue, research is underway to allow humans to task, train, and supervise robots through intuitive channels of communication. The goal is to design the interface such that untrained users can make safe and efficient use of the robot. Classically, these include the use of natural speech and gesture to task, train, and supervise the robot. The design challenge is to endow the robot with enough autonomy to carry out its assigned task while being able to interpret and respond to commands issued by the human overseeing its performance (Wilkes, Alford, Pack, Rogers, Peters & Kawamura 1997). The human must receive enough feedback (camera images, microphones, and force feedback) to be able to supervise the robot accordingly. Klingspor raises the issue of human-robot communication and discusses a scheme for how robots could learn a lexicon (a mapping from its sensory perceptions to linguistic symbols). This would allow a human to gradually teach the robot a more sophisticated instruction set, which would allow the robot to be issued

*Support for this research was provided by a MURI grant under the Office of Naval Research contract N00014–95–1–0600 and the Santa Fe Institute.

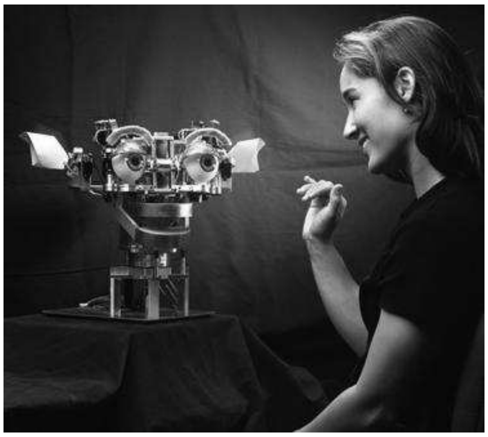

Figure 1: Kismet engaging a human caregiver using gaze direction and facial expressions.

more complex tasks over time (Klingspor et al. 1997).

3 Socially Intelligent Robots

The above approach views robots as very sophisticated appliances that people use to perform tasks. Some applications, however, require a more social form human-robot interaction. For these applications, an important part of the robot's functionality and usability is its ability to engage people in natural social exchange. Interactive computer-animated agents are typical modern-day precursors. Examples include animated characters such as *Silas*[1], personal digital assistants such as Microsoft's *Peety*[2], and discourse systems such as *Gandalf*[3].

Through building these types of systems researchers have found that as the amount of interactivity increases, people want these characters to be more *believable*[4] (Bates 1994). Classical and computer animation are full of examples where people are willing to "suspend disbelief" in order to interpret the character's behavior in human and social terms. Similarly, people shamelessly anthropomorphize their pets, computers, toys, etc., assigning them intentional, mental and emotional states (Watt 1995). By doing so, the entity seems more familiar and understandable to the human which in turn makes the interaction more comfortable, enjoyable, and compelling.

4 A Question of Interface

Within the field of cognitive technology Marsh, Nehaniv & Gorayska (1997) outline four key issues to bridge the gap between humans and interactive technologies. Dautenhahn (1998) refines these issues to propose several design considerations for socially intelligent agents (SIAs). For the purposes of this paper we concentrate on the following four design issues. First, how do people perceive SIARs, and how do these perceptions influence the way people interact with these technologies? Because *human perceptions of SIARs* shapes their expectations for how the technology should behave, users will consequently judge SIAR performance accordingly. Second, when interacting with SIARs, what channels of communication are the most natural, and how do these modalities shape the interaction? The *natural communication* issue determines which communication modalities constitute a natural and intuitive interface. Third, how do interactions with SIARS impact people on an emotional level? The *affective impact* issue becomes increasingly important as people integrate these technologies into their personal lives. And last, designers must consider the *social constraints* that shape the nature and quality of the interaction that people have (and expect to have) with SIARs.

Extending these findings to the robotics domain suggests that people will prefer SIARs to behave as socially intelligent creatures. To achieve this goal, the design of SIARs should address the following four human-interface issues: 1) human perception of SIARs, 2) natural communication, 3) affective impacts, and 4) social constraints.

4.1 Human Perception of SIARs

Humans are intentional creatures and perceive those they interact with as intentional creatures. Dennett's *intentional stance* argues that people explain their own behavior and interpret that of others in terms of intentions, beliefs, and desires (Dennett 1987). Hence, the ability of SIARs to convey intentionality to the user is an important design consideration. This doesn't require endowing machines with intents, beliefs, and wishes in the human sense, but the user should be able to intuitively and reliably explain and predict the robot's behavior in these terms.

Classical animators are masters at conveying intentionality through characters. In the *Illusion of Life*, Thomas and Johnston stress the importance of emotive expression for making animated characters believable (Thomas & Johnston 1981). They argue that it is *how* characters express themselves that conveys apparent beliefs, intents, and desires to the human observer. Reilly has applied these concepts using a "shallow but broad" approach to build interactive computer animated characters that are expressive and interact socially with each other (Reilly 1996). However, it is doubtful that superficial mechanisms will scale to unconstrained social interactions between humans and SIARs. Eventually, the expressive acts of SIARs may need to be generated by the equivalent of synthetic emotions.

4.2 Natural Communication with SIARs

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 informa-

[1] An animated dog of the *ALIVE* project (Blumberg 1996).

[2] An animated bird that users can query to play songs.

[3] An animated interface that users can query about the solar system during face to face exchange (Thorisson 1998).

[4] Some aspect of the character's behavior must appear natural, appealing, and life-like.

tion 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, etc.) as well as social cues (e.g., direction of gaze, feedback gestures such as nods of the head, raising eyebrows, etc.) to infer the intents, beliefs, and wishes of the other. Furthermore, people use these same expressive and social cues to *regulate the rate and content* of information transferred. For instance, people slow down or repeat themselves if the listener looks confused or uncertain. In this way, information is exchanged at a rate that is appropriate for both parties, and the sender can determine whether her message was received as intended.

To engage humans in natural social exchange, SIARs must be able to *read* the affective and social cues coming from its user. This entails visually perceiving the user's face, body, and eyes for affect, gesture, and gaze direction. Perceiving vocal prosody for intent and affect in addition to recognizing the user's speech is important as well. In addition, SIARs must also be able to *send* these same cues back to the user. This strongly argues for endowing SIARs with expressive faces, eyes that shift gaze, voices with variable prosody, and bodies that can posture expressively. By doing so, both parties will have access to the relevant information needed to communicate on factual and motivational levels, as well as to regulate the rate of information exchange.

4.3 Affective Impact of SIARs

As SIARs become more expressive and adept at social interaction, they will undoubtedly have a strong affective impact on their users. People freely anthropomorphize their pets, empathize with them, care about them, and emotionally bond to them. This process is facilitated when people perceive their pets as responding to them on an emotional level (wagging tails in happiness, howling in loneliness, growling in anger, etc.). People respond similarly to believable characters that masterfully portray powerful emotional responses to other characters and events. And although their expressive abilities are more limited, interactive computer agents such as petz or zorns, or interactive toys such as furbies or tamagotchis appeal to these same basic human tendencies. Hence when SIARs are able to expressively respond to their users, people will be naturally inclined to believe that a personal and emotional connection has been established between them. As a result, users will be likely to empathize with, care for, and emotionally bond to these technologies.

However, SIAR research is a long way to go before robots are establishing emotional ties with humans. Until then, a more practical concern is how to design SIARs so that they do not have a negative affective impact on users. Users are likely to find SIARs annoying or frustrating when SIARs fail to meet their expectations. Much of these expectations are shaped by the SIARs appearance. For instance, users will expect more mature and human-level interactions from an SIAR that resembles an adult human, than from one that resembles a cartoonish bird. Hence, care must be taken when designing SIARs to not implicitly set up false expectations, especially at such an early stage in SIAR research. Successful designs will poise the user to naturally interact with the robot at the right level of sophistication.

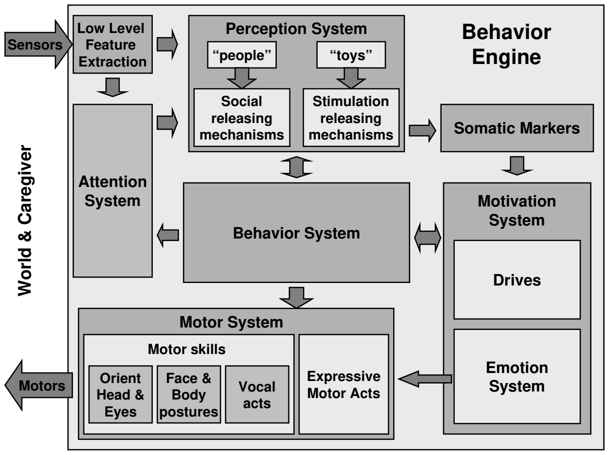

Figure 2: Overview of the software architecture. Perception, attention, internal drives, emotions, and motor skills are integrated to provide rich social interactions.

4.4 Human Social Constraints

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 interaction with others in the social world versus interaction 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 manner in which humans can *predict and socially influence* the behavior of the other through communicative acts. Much of this predictive power relies on each party being cooperative, open to communication, and subject to social norms. Given such, the ability of each person to assume the intentional stance and to empathize with the other helps each to predict and explain the behavior of the other, and to formulate appropriate responses based on this understanding.

To enable SIARs to engage in human-style social interaction, it may be necessary to endow them with a mechanism for empathy. Such a mechanism could enable SIARs to predict and explain its user's behavior, to properly formulate responses, to share control, and to mutually regulate the interaction. Furthermore, as SIARs become prevalent throughout society, users will need some way of culturally integrating them. An empathetic mechanism may also be necessary for teaching SIARs the *meaning* of their actions to others, and *value* of their actions to society. This goes far beyond learning a mapping from sensory readings to a lexicon of symbols. Somehow the agent must be able to evaluate the *goodness* versus *badness* of its actions with respect to the resulting consequences. Without this sense of value, the agent will not be able to discern socially acceptable actions from unacceptable ones, especially when a given action may be socially acceptable in one context, but highly inappropriate in another. Some of this commonsense social knowledge can be programmed in at design time, but ultimately learning will play an important role, especially if the robot is able to learn new behaviors.

4.5 Synthetic Emotions for SIARs

In each of the design considerations outlined above, emotions play a critical role in human-style social interaction. In highly constrained scenarios, the designer may be able to achieve believable interactions by heavily relying on the human's natural bias to attribute intentionality to the robot and to interact with it as if it were an intentional creature. However, to partake in human-style interaction in unconstrained social scenarios, SIARs must be able to adeptly express affective states to the human, perceive and interpret the human's emotive expressions, and use this information to determine appropriate social responses and to learn of their social consequences. Hence, incorporating synthetic emotions, empathetic learning mechanisms, and affective reasoning abilities into SIAR control architectures may be a critical step for reaching this level of believability in performance.

5 Architecture of an SIAR

To explore issues in building SIARs, we have been developing a robot called Kismet (see figure 1). Kismet has an active stereo color vision system, stereo microphones, a speech synthesizer, a two degree of freedom neck, and facial features which enable it to display a wide variety of recognizable expressions. Kismet is situated in a benevolent and social environment. Its task is to engage people in face to face social interaction and to improve its social competence from these exchanges. The interaction scenario is that of a robot infant playing with its human caregiver. Accordingly, Kismet has a highly expressive face with an infant-like appearance. This encourages people to naturally interact with Kismet as if it were a baby (approx. 6 months) and to teach it social skills commensurate with that age.

For the remainder of this paper, we present Kismet's control architecture and highlight how affective influences pervade the system. The architecture integrates perception, attention, internal drives, emotions[5], and motor skills (see Figure 2). Kismet's emotion processes

[5]When speaking of Kismet's "emotions", we are referring to computational processes. We do not claim to be implementing emotions in the human sense.

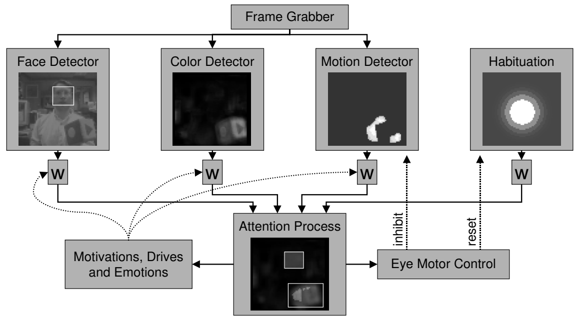

Figure 3: Kismet's attention and perception systems. Low-level feature detectors are combined with top-down motivational influences and habituation effects to direct eye and neck movements. In these images, Kismet has identified two salient objects: a face and a colorful toy block.

play an important role in biasing attention and behavior, and will ultimately play a critical role in socially situated learning (this is work in progress). The emotion processes also influence the expressive motor acts of Kismet through facial expressions, and eventually through voice and posture as well. In turn, these emotion processes are influenced by perception, motivations, and behavior. We illustrate how this architecture enables the robot to engage a human in infant-like social exchange.

6 Attention and Perception

Human infants show a preference for stimuli that exhibit certain low-level feature properties. For example, a four-month-old infant is more likely to look at a moving object than a static one, or prefer a face over other pleasing stimuli (Trevarthen 1979). To mimic the preferences of human infants, Kismet's perceptual system combines three basic feature detectors: face finding, motion detection, and color saliency analysis at speeds that are amenable to social interaction (20-30Hz).

Low-level perceptual inputs are combined with high-level influences from motivations and habituation effects by the attention system (see Figure 3). This system is based upon models of adult human visual search and attention (Wolfe 1994), and has been reported previously (Breazeal & Scassellati 1999). The attention process constructs a linear combination of the input feature detectors and a time-decayed Gaussian field which represents habituation effects. High areas of activation in this composite generate a saccade to that location and compensatory neck movement. The weights of the feature detectors can be influenced by the motivational (drives and emotions) and behavioral state of the robot to preferentially bias certain stimuli. For example, if the robot has become "lonely", the weight of the face detector can be increased to cause the robot to show a preference for attending to faces.

Perceptual stimuli selected by the attention process

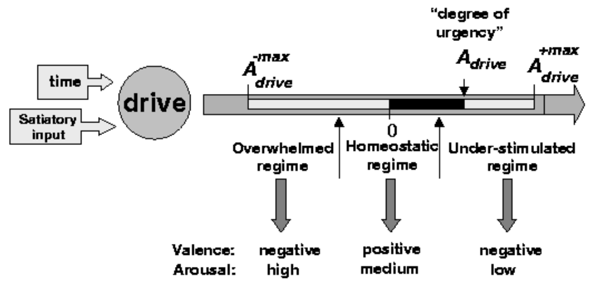

Figure 4: Kismet's affective state can be influenced by the drives. The overwhelmed regime has an agitating effect, the homeostatic regime has a calming effect, and the under-stimulated regime has a depressing effect.

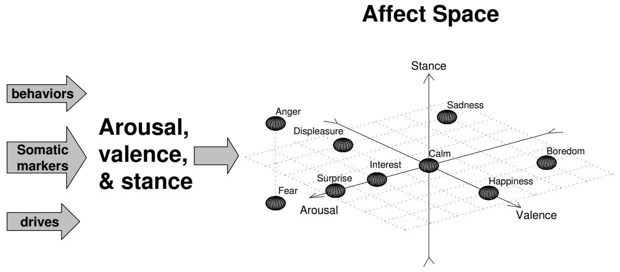

Figure 5: Kismet's affective state can be represented as a point along three dimensions: arousal, valence, and stance. This affect space is divided into `emotion` regions whose centers are shown here.

are classified into *social* stimuli (i.e., people, which move and have faces) which satisfy a drive to be social, and *non-social* stimuli (i.e., toys, which move and are colorful) which satisfy a drive to be stimulated by other things in the environment. This distinction can be observed in infants through a preferential looking paradigm (Trevarthen 1979).

The percepts for a given classification are then combined into a set of *releasing mechanisms.* Releasing mechanisms are used to specify both the nature of the stimulus (social vs non-social) and its quality (presence, absence, intensity, centered within the visual field, etc.). As conceptualized by ethologists, releasing mechanisms encapsulate the minimal environmental pre-conditions necessary for a behavior to become active (Lorenz 1973), (Tinbergen 1951).

Each releasing mechanism is tagged with arousal, valence, and stance values by an associated *somatic marker* process. This technique is inspired by the *Somatic Marker Hypothesis* of Damasio (1994) and enables percepts to influence the affective state of the robot through the releasing mechanisms. In this way, a good quality stimulus can make the robot's valence measure more positive (ultimately making the robot appear more pleased), an absence of stimuli may result in a decrease of arousal (making the robot appear bored or disinterested), and an overly intense stimuli may cause the robot's stance to become more closed (making the robot appear aversive towards the stimulus). In this way, Kismet has a variety of affective responses towards faces or colorful toys when interacting with them.

7 Motivations

Kismet's motivation system consists of drives and emotions. Drives represent the basic "needs" of the robot: a need to interact with people (the `social` drive), a need to be stimulated by toys and other objects (the `stimulation` drive), and a need for rest (the `fatigue` drive). For each drive, there is a desired operation point and an acceptable bounds of operation around that point (the homeostatic regime). Unattended, drives drift toward an under-stimulated regime. Excessive stimulation (too many stimuli or stimuli moving too quickly) push a drive toward an over-stimulated regime. When the intensity level of the drive leaves the homeostatic regime, the robot becomes motivated to act in ways that will restore the drives to the homeostatic regime.

Each drive influences the robot's affective state by contributing to the valence and arousal measures (see figure 4). When a drive is in the homeostatic regime, it adjusts the arousal and valence levels toward mid-range values, having a calming effect on the robot. When a drive is in the under-whelmed regime, it contributes to lower arousal and more negative valence, having a depressing effect on the robot. When a drive is in the over-whelmed regime, it contributes to higher arousal and more negative valence, having an agitating effect on the robot. By doing so, the robot's expression is commensurate with its current state of well being.

The robot's emotions are a result of its affective state. The affective state of the robot is represented as a point along three dimensions: *arousal* (i.e. high, neutral, or low), *valence* (i.e. positive, neutral, or negative), and *stance* (i.e. open, neutral, or closed). This space is termed the *affect space* as shown in figure 5. Russell (1980) presents a similar scheme (based on arousal, valence, and potency) to categorize emotions in people. The robot's current affective state is computed by summing contributions from the drives, somatically marked releasing mechanisms, and behaviors. The advantage of this representational scheme is that Kismet is in a distinct affective state at any one time, and the state varies smoothly within this space making for sensible transitions between different states.

To influence behavior, the affect space is compartmentalized into a set of emotion regions. Each region is characteristic of a particular emotion in humans. For example, `happiness` is characterized by positive valence, neutral arousal and neutral stance, whereas `sadness` is characterized as negative valence, low arousal, and neutral stance. The region whose center is closest to the current affect state is considered to be active. The intensity of the emotion is proportional to the radial distance from the origin to the current point in affect space.

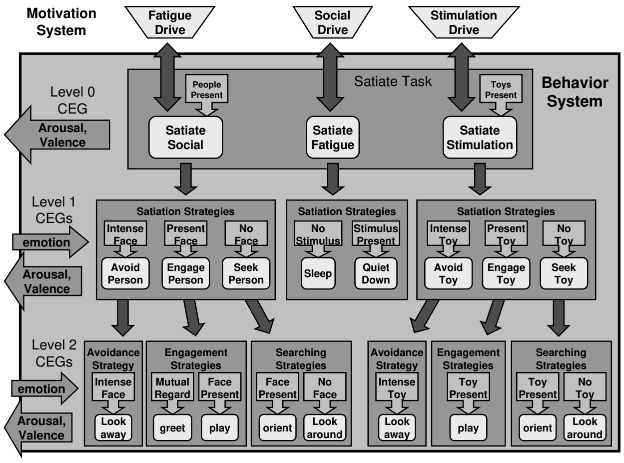

Figure 6: Kismet's behavior hierarchy consists of three levels of behaviors. Top level behaviors connect directly to drives, and bottom-level behaviors produce motor responses. Cross exclusion groups (CEG) conduct winner-take-all competitions to allow only one behavior in the group to be active at a given time.

Each emotion region has an affiliated emotion response process, implemented in a manner similar to that of (Velasquez 1997).

The motivational system influences behavior selection and the attentional focus based upon the current active emotion. The more intense the active emotion, the more strongly it influences these systems. As described previously, the motivational system influences the gains used to weight the relative contributions of each feature to the overall saliency of the stimulus. By doing so, more heavily weighed features become more salient depending on the robot's behavioral and motivational context. The motivations also pass activation to those behaviors that serve to restore the motivations to balance. For example, when in the `sadness` region of the affect space and in the `lonely` (i.e., under-stimulated) regime of the `social` drive, the motivational system applies a positive bias to behaviors that seek out people.

8 Behavior

Kismet's behavior system 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. 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 releasing mechanisms) to determine which behaviors become active. Once active, a behavior can influence both what the robot does (by influencing motor acts) and how that action is expressed through current facial expression (by influencing the arousal and valence aspects of the emotion system). Many of Kismet's behaviors are motivated by emotions as proposed by Plutchik (1984).

Behaviors are organized into a loosely layered, heterogeneous hierarchy as discussed in (Blumberg 1996). At each level, behaviors are grouped into *cross exclusion groups* (CEGs) which represent competing strategies for satisfying the goal of the parent; akin to "the avalanche effect" as discussed in (Minsky 1988). Within a CEG, a winner-take-all competition based on the current state of the emotions, drives, and percepts is held. The winning behavior may pass activation to its children (level 0 and 1 behaviors) or activate motor skill behaviors (level 2 behaviors).

Winning behaviors influence the current affective state, biasing towards a positive valence when the behavior is being applied successfully and towards a negative valence when the behavior is unsuccessful. In this way, the robot displays pleasure upon success, and growing frustration the longer it takes the active behavior to achieve its goal. Goals are often represented in perceptual terms. For instance, the goal for the `seek person` behavior is to have a face stimulus appear within the field of view. Until this event occurs, the robot engages in a visual search behavior.

Competition between behaviors at the top level (level 0) represents selection at the *global task* level. Level 0 behaviors receive activation based on the strength of their associated drive. Because the satiating stimuli for each drive are mutually exclusive and require different behaviors, all level 0 behaviors are members of a single CEG. This ensures that the robot can only act to restore one drive at a time. They also serve as a mechanisms for passing the arousal and valence influences of the drives to the emotion system. Only when a behavior is active can its affiliated drive pass its arousal and valence influences to the emotion system. In this way, only the drive currently being serviced can influence the expression on the robot's face, making it easier for a human to read and infer the robot's motivational state.

Competition between behaviors within the active level 1 CEG represents *strategy* decisions. Each level 1 behavior has its own distinct winning conditions based on the current state of the percepts, drives, and emotions. For example, the `avoid person` behavior is the most relevant when the `social` drive is in the overwhelmed regime and a person is stimulating the robot too vigorously. Similarly, `seek person` is relevant when the `social` drive is in the under-stimulated regime and no face percept is present. The `engage person` behavior is relevant when the `social` drive is already in the homeostatic regime and the robot is receiving a good quality stimulus. To preferentially bias the robot's attention to behaviorally relevant stimuli, the active level 1 behavior can adjust the feature gains of the attention system.

Competition between level 2 behaviors represents *subtask* divisions. For example, when the `seek person` behavior is active at level 1, if the robot can see a face then the `orient to face` behavior is activated. Otherwise, the `look around` behavior is active. Once the

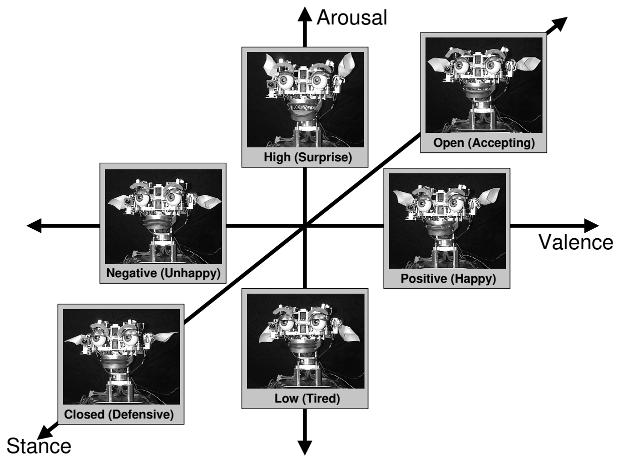

Figure 7: The expression space of facial postures for each dimension: high arousal, low arousal, positive valence, negative valence, open stance, and closed stance.

robot orients to a face, bringing it into mutual regard, the `engage person` behavior at level 1 becomes active. The `engage person` behavior activates its child CEG at level 2. The `greet` behavior becomes immediately active since the robot and human are in mutual regard. After the greeting is delivered, the internal persistence of the `greet` behavior decays and allows the `play` behavior to become active. Once the satiatory stimulus (in this case a face in mutual regard) has been obtained, the appropriate drive is adjusted according to the quality of the stimulus.

9 Expressive Motor Acts

The motor system receives input from the emotion subsystem and the behavior system. Level 2 behaviors evoke motor acts such as `look around` which moves the eyes to obtain a new visual scene, `look away` which moves the eyes and neck to avoid a noxious stimulus, `greet` which wiggles the ears while fixating on a persons face, and `orient` which produces a neck movement with compensatory eye movement to place an object in mutual regard.

Each dimension of the affect space (arousal, valence, and stance) is mapped to an *expression space* where each dimension has a characteristic facial posture for each extreme (see figure 7). Hence, Kismet has six prototypical expressions (the basis set) for high arousal, low arousal, negative valence, positive valence, open stance, and closed stance. These six facial postures span the space of all possible expressions Kismet can generate.

Although some dimensions adjust specific facial features more strongly than other dimensions, each dimension influences most if not all the facial features to some degree. Hence, valence has the strongest influence on lip curvature, but can also adjust the positions of the ears, eyelids, eyebrows, and jaw.

The basis set of facial expressions has been designed so that a specific location in affect space maps to a mutually consistent emotion process and facial expression. With this scheme, Kismet can display expressions analogous to anger, boredom, displeasure, fear, happiness, interest, sadness, surprise, calm, and a variety of others. The advantage of the expression space representation is that it allows Kismet to display a distinct and easily readable expression consistent with its affective state.

10 Social Learning

In previous work, we outline an approach (for work in progress) consisting of three learning mechanisms by which the caregiver could train Kismet using emotive channels of communication (e.g., facial expression or affective cues in voice) (Breazeal & Velasquez 1998). The caregiver does so by exploiting the learning mechanics of Kismet's motivation system, to place Kismet in either a positive affective state (positive reinforcement) when the robot does something desirable, or a negative affective state (negative reinforcement) when the robot does something undesirable. By doing so, Kismet's affective states mirror those of the caregiver. These learning mechanisms bias the robot to learn and pursue behaviors that please the caregiver and to avoid those that displease her. By communicating reward and punishment information through emotive channels, the caregiver can actively help Kismet *identify* and *pursue* new behaviors as they play together, by assigning them values of *goodness* (i.e., pleases caregiver) and *badness* (i.e., displeases caregiver). As such, these mechanisms could serve as a simple form of empathy and as a starting point for teaching Kismet the value of its actions to others. These sorts of social learning mechanisms are the focus of ongoing development.

11 Social Interaction

This architecture produces interaction dynamics similar to the five phases of infant-caregiver social interactions (*initiation*, *mutual-orientation*, *greeting*, *play-dialog*, and *disengagement*) described in Tronick, Als & Adamson (1979). 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 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.

Figure 8 shows Kismet's dynamic responses during face-to-face interaction with a caregiver. Kismet is initially looking for a person and displaying sadness (the initiation phase). The robot begins moving its eyes looking for a face stimulus ($t < 8$). When it finds the caregiver's face, it makes a large eye movement to enter into mutual regard ($t \approx 10$). Once the face is foveated, the

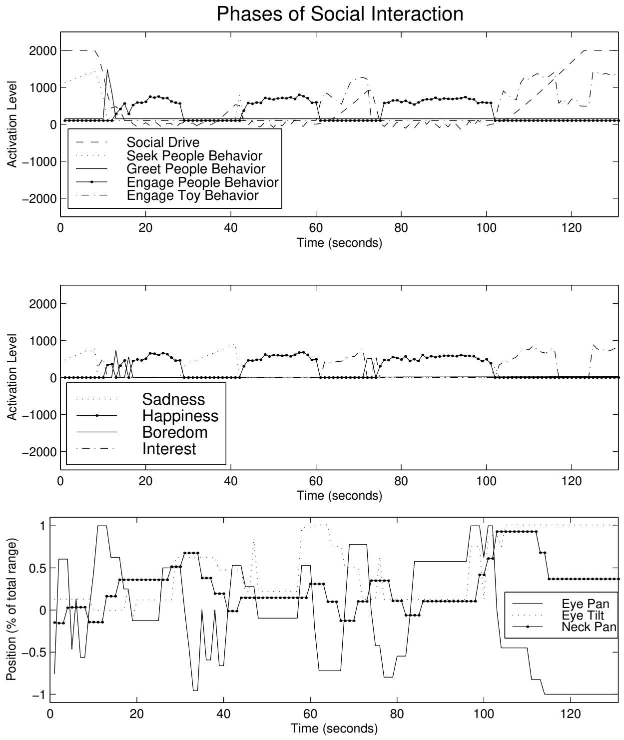

Figure 8: Cyclic responses during social interaction. Behaviors and drives (top), emotions (middle), and motor output (bottom) are plotted for a single trial of approximately 130 seconds. See text for description.

robot displays a greeting behavior by wiggling its ears ($t \approx 11$), and begins a play-dialog phase of interaction with the caregiver ($t > 12$). Kismet continues to engage the caregiver until the caregiver moves outside the field of view ($t \approx 28$). Kismet quickly becomes sad, and begins to search for a face, which it re-acquires when the caretaker returns ($t \approx 42$). Eventually, the robot habituates to the interaction with the caregiver and begins to attend to a toy that the caregiver has provided ($60 < t < 75$). While interacting with the toy, the robot displays interest and moves its eyes to follow the moving toy. Kismet soon habituates to this stimulus, and returns to its play-dialog with the caregiver ($75 < t < 100$). A final disengagement phase occurs ($t \approx 100$) when the robot's attention shifts back to the toy.

12 Summary

Our findings illustrate how the expression of synthetic emotions through overt behavior has the power to influence the social world. Kismet is able to successfully negotiate the caregiver into presenting the robot with toys when it is bored, and to engage in face to face exchange when it is lonely. All the while, the caregiver instinctively responds to the robot's affective state to promote its well being – presenting the robot with pleasing stimuli, avoiding the presentation of noxious stimuli, and taking care not to overwhelm nor under-stimulate the robot.

Much of Kismet's behavior has been designed to address the four human-robot interface issues highlighted in the first half of the paper. Specifically, 1) Kismet's goal directed behavior and expressive abilities enables the robot to convey intentionality to the caregiver; 2) its facial displays and directed gaze serve as readily interpretable social cues that the caregiver instinctively uses to support natural modes of communication; 3) its ever-changing affective displays in response to inter-personal exhanges with people has an affective impact upon the caregiver's behavior, eliciting nurturing responses from her to maintain the robot in a state of well being; and 4) through these social exchanges, the robot is able to share control with the caregiver in order to regulate the intensity of interaction and to encourage the caregiver to address the robot's basic "needs" throughout the exchange.

In doing so, we have shown how emotion-inspired mechanisms have a pervasive influence on the internal dynamics of the robot's controller, biasing perception, attention, motivation, behavior, learning, and the expression of motor acts. Whereas past approaches have focused on perception and task-based behavior, our approach balances these with affective factors and their expression. We believe this to be a critical step towards the design of socially intelligent synthetic creatures, which we may ultimately be able to interact with as friends instead of as appliances.

References

Bates, J. (1994), 'The role of emotion in believable characters', *Communications of the ACM* **37**(7), 122–125.

Blumberg, B. (1996), Old Tricks, New Dogs: Ethology and Interactive Creatures, PhD thesis, MIT.

Breazeal, C. & Scassellati, B. (1999), A context-dependent attention system for a social robot, *in* '1999 International Joint Conference on Artificial Intelligence'. Submitted.

Breazeal, C. & Velasquez, J. (1998), Toward teaching a robot "infant" using emotive communication acts, *in* 'Socially Situated Intelligence: Papers from the 1998 Simulated Adaptive Behavior Workshop'.

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

Dautenhahn, K. (1998), 'The art of designing socially intelligent agents: Science, fiction, and the human in the loop', *Applied Artificial Intelligence Journal* **12**(7–8), 573–617.

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

Inoue, H. (1998), A Platform-based Humanoid Robotics Project, *in* 'First International Workshop on Humanoid and Human Friendly Robots', Tsukuba, Japan, pp. I–1.

Klingspor, V., Demiris, J. & Kaiser, M. (1997), 'Human-Robot-Communication and Machine Learning', *Applied Artificial Intelligence Journal* **11**, 719–746.

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

Marsh, J., Nehaniv, C. & Gorayska, B. (1997), Cognitive Technology: Humanizing the information age, *in* J. Marsh, C. Nehaniv & B. Gorayska, eds, 'Second International Conference on Cognitive Technology', IEEE Computer Society Press, pp. vii–ix.

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

Plutchik, R. (1984), Emotions: A general psychevolutionary theory, *in* K. Scherer & P. Elkman, eds, 'Approaches to Emotion', Lawrence Erlbaum Associates, New Jersey, pp. 197–219.

Reilly, S. (1996), Believable Social and Emotional Agents, PhD thesis, CMU School of Computer Science.

Russell, J. (1980), 'A circumplex model of affect', *Journal of Personality and social psychology*.

Thomas, F. & Johnston, O. (1981), *Disney animation: The Illusion of Life*, Abbeville Press, New York, NY.

Thorisson, K. (1998), Real-time Decision Making in Multimodal Face-to-face Communication, *in* 'Second International Conference on Autonomous Agents', ACM SIGART, ACM Press, Minneapolis, MN, pp. 16–23.

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

Trevarthen, C. (1979), Communication and cooperation in early infancy: a description of primary intersubjectivity, *in* M. Bullowa, ed., 'Before Speech', Cambridge University Press, pp. 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, pp. 349–370.

Velasquez, J. (1997), Modeling Emotions and other motivations in synthetic agents, *in* 'Proceedings of the 1997 National Conference on Artificial Intelligence, AAAI97', pp. 10–15.

Watt, S. (1995), The naive psychology manifesto, Technical Report KMI-TR-12, The Open University, Knowledge Media Institute.

Wilkes, D., Alford, A., Pack, R., Rogers, R., Peters, R. & Kawamura, K. (1997), 'Toward Socially Intelligent Service Robots', *Applied Artificial Intelligence Journal* **12**, 729–766.

Wolfe, J. (1994), 'Guided Search 2.0: A revised model of visual search', *Psychonomic Bulletin and Review* **1**(2), 202–238.