Function meets Style: Insights from Emotion Theory Applied to HRI

Cynthia Breazeal

[1]

Abstract— **As robot designers, we tend to emphasize the cognitive aspect of intelligence when designing robot architectures while viewing the affective aspect with skepticism. However, scientific studies continue to reveal the deeply intertwined and complementary roles that cognition and emotion play in intelligent decision making, planning, learning, attention, communication, social interaction, memory, and more. Such findings provide valuable insights and lessons for the design of autonomous robots that must operate in complex and uncertain environments and perform in cooperation with people. This paper presents a concrete implementation of how these insights have guided our work, focusing on the design of sociable autonomous robots that interact with people as capable partners.**

Keywords— **human-robot interaction, humanoid robots, models of emotion and affect, social or sociable robots**

I. INTRODUCTION

All intelligent creatures that we know of have emotions. Humans, in particular, are the most expressive, emotionally complex, and socially sophisticated of all [1].

To function and survive in a complex and unpredictable environment, animals (including humans) were faced with applying their limited resources (e.g., muscles, limbs, perceptual systems, mental abilities, etc.) to realize multiple goals in an intelligent and flexible manner [2]. Those species considered to be the most intelligent tend to exist in complex and dynamic social groups where members have to communicate, cooperate, or compete with others. Two distinct and complementary information processing systems, cognition and emotion, evolved under such social and environmental pressures to promote the health and optimal functioning of the creature.

As argued by Norman, Ortony and Russell [3], these two systems are deeply intertwined and operate simultaneously. The cognitive system is responsible for interpreting and making sense of the world, whereas the emotion system is responsible for evaluating and judging events to assess their overall value with respect to the creature (e.g., positive or negative, desirable or undesirable, hospitable or harmful, etc.). When operating in the proper balance, the emotion system modulates the operating parameters of the cognitive system and the body to improve the overall mental and physical performance of the creature. The scientific literature documents the beneficial effect of emotion on creative problem solving, attention, perception, memory retrieval, decision-making, learning, and more (see [4] for an overview). As argued by Damasio [5], too much emotion can hinder intelligent thought and behavior, however, too little emotion is even more problematic.

II. WHY GIVE ROBOTS EMOTION SYSTEMS?

Today's autonomous robots can certainly improve their ability to function in complex environments and to behave appropriately in partnership with people. Using the design of natural intelligences as a guide, a robot's cognitive system would enable it figure out what to do, whereas the emotion system would help it to do so more flexibly in complex and uncertain environments, as well as help the robot behave in a socially acceptable and effective manner with people. It is in this pragmatic spirit that we explore the role of emotion-like processes and capabilities in human-robot interaction.

This does not imply that these emotion-based or cognition-based mechanisms and capabilities must be in some way identical to those in natural systems. In particular, the question of whether or not robots could have *real* emotions is irrelevant to our purposes. We agree with Picard [6]: as we continue to design integrated systems for robots, that complement its cognitive capabilities with those regulatory, signaling, biasing, and other useful attention, value assessment, and prioritization mechanisms associated with emotion systems in living creatures, we will effectively be giving robots a system that serves the same useful functions, no matter what we call it.

III. EMOTION-RELATED SKILLS AND MECHANISMS IN HRI

In order to interact with others (whether it is a device, a robot, or even another person) it is essential to have a good conceptual model for how the other operates [7]. With such a model, it is possible to explain and predict what the other is about to do, its reasons for doing it, and how to elicit a desired behavior from it. As argued by Norman [7], the design of a technological artifact, whether it is a robot, a computer, or a teapot, can help a person form this model by "projecting a image of its operation," either through visual cues or continual feedback. Hence, there is a very practical side to developing robots that can effectively convey and communicate their internal state to people for cooperative tasks, even when the style of interaction is not social.

Our work focuses on developing sociable robots for envisioned applications where the robot interacts with a person as a partner. There are many important emotion-related skills and mechanisms that would be useful if not necessary for the success of the robot in these applications. Fur-

---

[1] Manuscript received XXXX; revised XXXX. The author is with the Media Lab, Massachusetts Institute of Technology (MIT), Cambridge, MA (e-mail: cynthiab@media.mit.edu). This work was funded by MIT Media Lab corporate sponsors of the Things That Think and Digital Life consortia.

thermore, if the internal design of the robot mirrors the characteristics of the emotion systems of living creatures, then emotive expressions are very appropriate and intuitively understood by humans across different cultures [8], [9]. These capabilities are listed below. In the following sections, we illustrate how the emotion system and cognitive system work together to implement these capabilities on our sociable robot.

- Intelligent behavior in a complex, unpredictable environment.
- The ability to sense and recognize emotion and affect in others.
- The ability to express affect and internal state to others.
- The ability to respond to humans with social adeptness and appropriateness.

A robot that cares for the elderly, for instance, should be able to respond appropriately in times when the patient is showing signs of distress or anxiety. It should be persuasive in ways that are sensitive to the person, such as helping to remind them when to take medication, without being annoying or upsetting. It would need to know when to contact a health professional when necessary.

Yet so many current technologies, such as animated computer agents, interact with us in a manner characteristic of emotionally impaired people. In the best cases they know what to do, but often lack the emotional intelligence to do it in a socially appropriate manner. As a result, they frustrate us and we quickly dismiss them even though they can be useful. Given that many exciting applications for autonomous robots in the future place them in a long-term relationship with a particular person, we will need to address these issues or people will not accept them into their daily lives.

IV. Design of a Sociable Robot

The *Sociable Robots Project* develops expressive anthropomorphic robots to explore scientific questions and to address engineering challenges of building socially and emotionally intelligent robots (see Figure 1). Their social and emotive qualities are integrated deep into the core of their design, and serve not only to "lubricate" the interface between itself and its human interlocutor, but also play a pragmatic role in promoting survival, self maintenance, learning, decision making, attention, and more [10], [11]. Hence, social and affective interactions with people are valued not just at the interface, but at a pragmatic and functional level for the robot as well.

Inspired by models of intelligence in natural systems, the design of our architecture features both a cognitive system and an emotion system (see Figure 2). Both operate in parallel and are deeply intertwined to foster optimal functioning of the robot in its environment. The overall architecture is comprised of a distributed network of interacting agent-like processes that excite and inhibit one another by spreading activation [12]. Due to space constraints, we cannot give an equal presentation of both systems. Hence, we focus on the design of the emotion system and provide an overview of the cognitive system. We use our implementation on our robot, Kismet, as a case study and refer the reader to [10] for greater detail. These ideas are being extended to explore socially mediated learning on our current robot, Leonardo, the successor of Kismet (see Figure 1).

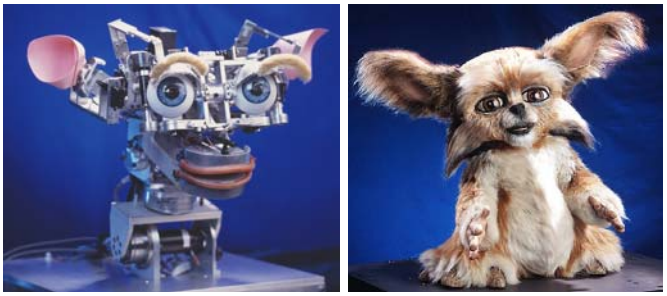

Fig. 1. Kismet(left) and Leonardo (right), our sociable robots. Whereas Kismet has a mechanical aesthetic, Leonardo has a more organic appearance. Photos courtesy of Sam Ogden.

V. The Cognitive System

The cognitive system is responsible for perceiving and interpreting events, and arbitrating among the robot's goal-achieving behaviors to address competing motives. Each motive is modeled as a homeostatic process that represents the robot's "health" related goals. The computational systems and mechanisms that comprise the cognitive system work in concert to decide which behavior to activate at what time and for how long to service the appropriate objective. Overall, the robot's behavior must exhibit an appropriate degree of relevance, persistence, flexibility, and robustness. To achieve this, we based the design of the cognitive system on ethological models of animal behavior [2].

A. Perceptual Elicitors

Inputs arising from the environment originate from the perceptual system where key features are extracted from the robot's sensors (cameras, microphones, etc.). These features are fed into an associated *releaser* process. Each releaser can be thought of as a simple perceptual elicitor of behavior that combines lower-level features into behaviorally significant perceptual categories. For instance, the visual features of `color`, `size`, `motion`, `proximity` are integrated to form a `toy` percept. Other releasers are designed to characterize important internal events, such as the urgency to tend to a particular motive. There are many different releasers designed for Kismet, each signals a particular event or object of behavioral significance. If the input features are present and of sufficient intensity, the activation level of the releaser process rises above threshold, signifying that the conditions specified by that releaser hold. Given this, its output is passed to its corresponding behavior process in the behavior system, thereby preferentially contributing to that behavior's activation.

B. Motivations

Kismet's *drives* implement autopoetic-related processes for satisfying the robot's "health" related and time-varying

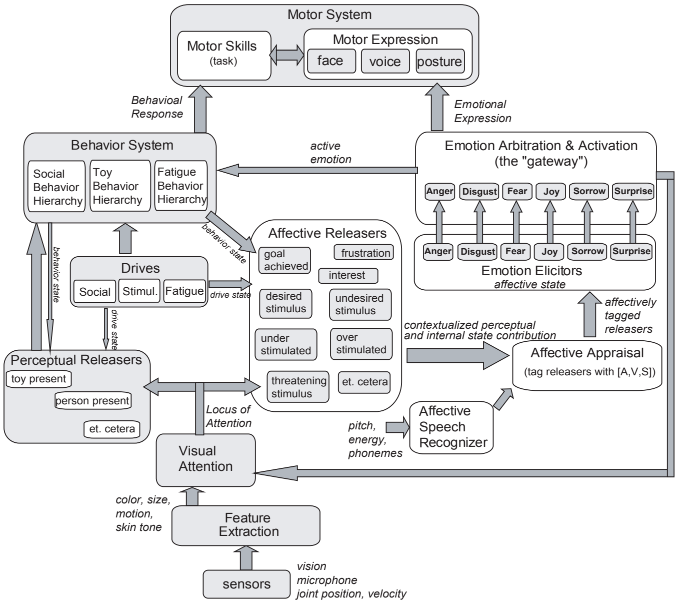

Fig. 2. An architectural overview showing the tight integration of the *cognitive system* (perception, attention, behavior), with the *emotion system* (affective appraisal, emotive elicitors, emotion "gateway" processes) shown as white modules, and the *motor system*.

goals. Analgous to the motivations of "thrist", "hunger", and "fatigue" for an animal, Kismet's drives motivate it to get the right amount of the desired kind of stimulation in a timely manner. Kismet's drives correspond to a "need" to interact with people (the `social-drive`), to be stimulated by toys (the `stimulation-drive`), and to occasionally rest (the `fatigue-drive`). The degree to which each drive is satiated in a timely fashion contributes to the robot's overall measure of its "well being."

The design of each drive is heavily inspired by ethological views of the analogous process in animals. Their change in intensity reflects the ongoing "needs" of the robot and the urgency for tending to them. Each drive acts to maintain a level of intensity within a bounded range, neither too much nor too little, as defined by a desired operational point and acceptable bounds of operation around that point (called the homeostatic regime). A drive remains in its homeostatic regime when it is encountering its satiatory stimulus of appropriate intensity. Given no satiatory stimulation, a drive will tend to increase in (positive) intensity.

Motivations shape the internal agenda of the robot and play an important role in determining which behavior to engage in next. To keep its activation level within the homeostatic regime, each drive can preferentially spread activation to behaviors that help to restore it. For instance, in the absence of the satiatory stimulus (or if the intensity is too low), the drive increases in intensity to the positive end of the spectrum and preferentially biases the activation of those behaviors that serve to seek out that stimulus. Alternatively, if the satiatory stimulus is too intense (e.g., moving too close or too fast), the drive tends toward the extreme negative end of the spectrum. In this circumstance, the drive biases the activation of avoidance behaviors to limit the robot's exposure to the intense stimulus. Hence, to remain in balance (near the center of the spectrum), it is not sufficient that the satiatory stimulus be present; it must also be of a good quality.

C. Behavior Arbitration as Decision Making

Within the behavior system, the behavior processes are organized into loosely layered, heterogeneous hierarchies of behavior groups, much in the spirit of those ethological models proposed by Tinbergen [13] and Lorenz [14]. Implicit in this model is that at every level of the hierarchy a decision is being made among several alternatives of which one is chosen. At the top, the decisions are very general (which drive to satiate) and become increasingly more specific as one moves down a hierarchy. At the highest level, behaviors are organized into competing *functional groups* (the primary branches of the hierarchy) where each group is responsible for maintaining one of the three homeostatic functions. Only one functional group can be active at a time. The influence of the drives is strongest at the top level of the hierarchy, biasing which functional group should be active.

Each functional group consists of an organized hierarchy of *behavior groups* (akin to Tinbergen's behavioral centers) [13]. At each level in the hierarchy, each behavior group represents a competing strategy (a collection of behaviors) for satisfying the goal of its parent behavior. In turn, each *behavior* within a behavior group is viewed as a task-achieving entity whose particular goal contributes to the strategy of its behavior group. Each behavior process within a group competes with the others in a winner-take-all fashion for expression based on its relevance to the current situation as determined by the perceived environment (through its releaser processes) and its motives (through its drive processes). When active, a behavior coordinates sensori-motor patterns to achieve a particular task such as search behaviors, approach behaviors, avoidance behaviors, and interaction behaviors.

The observed behavior of the robot is the result of competition at the functional, strategic, and task levels. At the behavioral category level, the functional groups compete to determine which "need" is to be met (for Kismet, this corresponds to socializing, playing, or sleeping). At the strategy level, behavior groups of the winning functional group compete for expression. Finally, on the task level, the behaviors of the winning behavior group compete for expression. As one moves down in depth, the behaviors serve to more finely tune the relation between the robot and its environment, and in particular, the relation between the robot and the human [10].

VI. The Emotion System

The emotion system is responsible for perceiving and recognizing internal and external events with affective value, assessing and signaling this value to other systems, regulating and biasing the cognitive system to promote appropriate and flexible decision making, and communicating the robot's internal state to others to socially regulate their behavior in a beneficial relation to the robot. In concert with the robot's cognitive system, it is designed to be a flexible system that mediates between both environmental and internal events to elicit an adaptive behavioral response that serves either social or self-maintenance functions.

The organization and operation of the emotion system is strongly inspired by various theories of *basic emotions* in humans [8]. These few select emotions are endowed by evolution because of their proven ability to facilitate adaptive responses that promote a creature's daily survival in a complex and often hostile environment. Supported by evolutionary, developmental, and cross-cultural studies, these basic emotions include anger, disgust, fear, joy, sorrow, and surprise. For Kismet, an emotional response consists of:

- A precipitating event.
- An affective appraisal of that event.
- A characteristic display that can be expressed through facial expression, vocal quality, or body posture.
- Modulation of the cognitive and motor systems to motivate a behavioral response.

A. *Affective Releasers*

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

A.1 Recognition of communicated affect

Objects that Kismet interacts with can have affective value, such as a toy that is moving in a threatening manner. However, people can communicate affect directly to Kismet through tone of voice [15]. Developmental psycholinguists can tell us quite a lot about how preverbal infants achieve this. Based on a series of cross-linguistic analyses, there appear to be at least four different pitch contours that prelinguistic infants can recognize affectively (i.e., approval, prohibition, comfort, and attentional bids), each associated with a different emotional state [16].

Inspired by these theories, we have implemented a recognizer for distinguishing these four distinct prosodic patterns from Kismet-directed speech. The implemented classifier consists of several mini-classifiers executing in stages. A very detailed presentation of the recognizer and its performance assessment can be found in [10].

In all training phases we modeled each class of data using the Gaussian mixture model, updated with the EM algorithm and a Kurtosis-based approach for dynamically deciding the appropriate number of kernels [17]. The idea of the Gaussian mixture model is to represent the distribution of a data vector by a weighted mixture of component models, each one parameterized on its own set of parameters. Formally, the mixture density for the vector x assuming k components is

$$p(x) = \sum_{j=1}^{k} \pi_j f(x;\phi_j)$$

where $f(x;\phi_j)$ is the j-th component model parameterized on ϕ_j, π_j are the mixing weights satisfying $\sum_{j=1}^{k} \pi_j = 1$, and $\pi_j \geq 0$.

In this algorithm, kurtosis is viewed as a measure of non-normality and is used to decide on the number of components in the Gaussian problem. For a random vector x with mean m and covariance matrix S, the weighted kurtosis is defined as

$$\beta j = \sum_{i=1}^{n} P(j|x_i) \frac{((x_i - m_j)^T S_j^{-1}(x_i - m_j))^2}{\sum_{i=1}^{n} P(j|x_i)}.$$

Iteratively, EM steps are applied until convergence, and a new component is added dynamically until the test of normality $B = [\beta - d(d+2)]/\sqrt{[8d(d+2)]/n}$ indicates that $|B| \leq T$ for a predefined threshold, T.

Based on our recordings, the preprocessed pitch contours from the training set resemble Fernald's prototypical prosodic contours for approval, attention, prohibition, comfort/soothing, and neutral. Hence, we used Fernald's insights to select those features that would prove useful in distinguishing these five classes.

For the first classifier stage, global pitch and energy features (i.e., pitch mean and energy variance) partitioned the samples into useful intermediate classes. For instance, the prohibition samples are clustered in the low pitch mean and high energy variance region. The approval and attention classes form a cluster at the high pitch mean and high energy variance region. The soothing samples are clustered in the low pitch mean and low energy variance region. Finally, the neutral samples have low pitch mean, but are divided into two regions in terms of their energy variance values. The structure of each of the mini-classifiers follows logically from these observations. Table I shows the resulting classification performance.

B. *Affective Appraisal*

In Kismet's implementation, there is an explicit assessment phase for each active releaser, of which there are a number of factors that contribute to the assessment made.

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

TABLE I

Overall classification performance evaluated using a new test set of 371 utterances from adult female speakers.

The assessment consists of labeling the releaser with affective tags, a mechanism inspired by Damasio's *somatic marker hypothesis*, where incoming perceptual, behavioral, or motivational information is "tagged" with affective information [5].

For example, there are three classes of tags used within Kismet to affectively characterize a given releaser. Each tag has an associated intensity that scales its contribution to the overall affective state. The arousal tag, A, specifies how arousing this factor is to the emotional system. It very roughly corresponds to the activity of the autonomic nervous system. Positive values correspond to a high arousal stimulus whereas negative values correspond to a low arousal stimulus. The valence tag, V, specifies how favorable or unfavorable this percept is to the robot. Positive values correspond to a pleasant stimulus whereas negative values correspond to an unpleasant stimulus. The stance tag, S, specifies how approachable the percept is. Positive values correspond to advance whereas negative values correspond to retreat.

There are three factors that contribute to an appraisal of an active releaser. The first is the intensity of the stimulus, which generally maps to arousal. Threatening or very intense stimuli are tagged with high arousal. Absent or low intensity stimuli are tagged with low arousal. The second is the relevance of the stimulus to whether it addresses the current goals of the robot. This influences the valence and stance values. Stimuli that are relevant are desirable and are tagged with positive valence and approaching stance. Stimuli that are not relevant are undesirable and are tagged with negative arousal and withdrawing stance. The third factor is intrinsic pleasantness. Some stimuli are hardwired to influence the robot's affective state in a specific manner. For instance, praising speech is tagged with positive valence and slightly high arousal, whereas scolding speech is tagged with negative valence and low arousal, which tends to elicit as sorrowful response. In Kismet, there is a fixed mapping from each of these factors to how much they contribute to arousal, valence, or stance.

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

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

C. Emotion Elicitors, Activation, and Arbitration

All somatically marked inputs (e.g., releasers, the state of each drive, etc.) are passed to the emotion elicitors. There is an elicitor associated with each basic emotion process (e.g., `anger`, `fear`, `disgust`, etc.). The elicitor determines the relevance of its emotive response based on the myriad of factors contributing to it. In a living creature, this might include neural factors, sensorimotor factors, motivational factors, and cognitive factors [18]. Each elicitor computes the relevance of its affiliated emotion process and contributes to its activation. Each elicitor can thus be modeled as a process that computes its activation energy, $E_{emot}(i)$, for emotion, i, according to the function,

$$\begin{aligned} E_{emot}(i) &= R_{emot}(i) + Dr_{emot}(i) + Em^{excite}_{emot}(i) \\ &\quad - Em^{inhibit}_{emot}(i) + Bh_{emot}(i) \end{aligned}$$

Given the following somatically marked factors: $R_{emot}(i)$ is the weighted contribution of the active releasers,

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

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

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

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

When active, each emotion process acts as a gateway that when "open" can spread activation to a number of different cognitive systems (i.e., behavior, attention, expression, etc.). As a result, the emotive state of the robot is distributed throughout the overall architecture. Each emotion process plays a distinct regulatory role, modulating these systems in a characteristic manner when active. In a process of *behavioral homeostasis*, the emotive response maintains activity through external and internal feedback until the correct relation of robot to environment is established [19]. Concurrently, the affective state of the robot, as specified by the net $[A, V, S]$ of the active process, is sent to the expressive components of the motor system, causing a distinct facial expression, vocal quality, and body posture to arise.

VII. Expression of Affective State

Kismet can communicate its emotive state and other social cues to a human through facial expressions, body posture, gaze direction, and quality of voice [20]. We do not have sufficient space to explain how each of these is implemented, however they all contribute to the readability of Kismet's expression. We have found that the scientific basis for how emotion correlates to facial expression or vocal expression to be very useful in mapping Kismet's emotive states to its face actuators and to its articulatory-based speech synthesizer, respectively.

With respect to communicating emotion through the face, psychologists of the *componential theory* of facial expression posit that these expressions have a systematic, coherent, and meaningful structure that can be mapped to affective dimensions that span the relationship between different emotions [21]. Some of the individual features of expression have inherent signal value. The raised brows, for instance, convey attentional activity for both the expression of fear and surprise. By considering the individual facial action components that contribute to the overall facial display, it is possible to infer much about the underlying properties of the emotion being expressed. This promotes a signaling system that is robust, flexible, and resilient. It allows for the mixing of these components to convey a wide range of affective messages, instead of being restricted to a fixed pattern for each emotion.

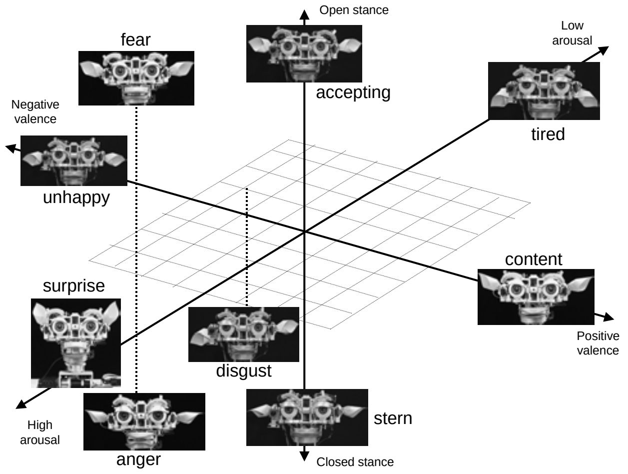

Fig. 3. This diagram illustrates where the basis postures are located in affect space.

Inspired by this theory, Kismet's facial expressions are generated using an interpolation-based technique over a three-dimensional affect space (see Figure 3). The three dimensions correspond to arousal, valence, and stance — the same three attributes that are used to affectively assess the myriad of environmental and internal factors that contribute to Kismet's overall affective state. There are nine *basis postures* that collectively span this space of emotive expressions.

The current affective state (as defined by the net $[A, V, S]$) occupies a single point in this space at a time. As the robot's affective state changes, this point moves around this space and the robot's facial expression changes to mirror this. As positive valence increases, Kismet's lips turn upward, the mouth opens, and the eyebrows relax. However, as valence decreases, the brows furrow, the jaw closes, and the lips turn downward. Along the arousal dimension, the ears perk, the eyes widen, and the mouth opens as arousal increases. Along the stance dimension, increasing positive values cause the eyebrows to arc outwards, the mouth to open, the ears to open, and the eyes to widen. The expressions become more intense as the affect state moves to more extreme values in the affect space. Hence, Kismet's face functions as a window by which a person can view the robot's underlying affective state. This transparency plays an important role in providing the human with the necessary feedback to understand and predict the robot's behavior.

Antecedent Conditions	*Emotion*	*Behavior*	*Function*
Delay, difficulty in achieving goal of adaptive behavior	anger, frustration	complain	show displeasure to caregiver to modify his/her behavior
Presence of an undesired stimulus	disgust	withdraw	signal rejection of presented stimulus to caregiver
Presence of a threatening, overwhelming stimulus	fear, distress	escape	Move away from a potentially dangerous stimuli
Prolonged presence of a desired stimulus	calm	engage	Continued interaction with a desired stimulus
Success in achieving goal of active behavior, or praise	joy	display pleasure	Reallocate resources to the next relevant behavior (eventually to reinforce behavior)
Prolonged absence of a desired stimulus, or prohibition	sorrow	display sorrow	Evoke sympathy and attention from caregiver (eventually to discourage behavior)
A sudden, close stimulus	suprise	startle response	alert
Appearance of a desired stimulus	interest	orient	attend to new, salient object
Need of an absent and desired stimulus	boredom	seek	Explore environment for desired stimulus

TABLE II

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

VIII. Integrated Emotive Responses

As shown in Table II, a number of emotive responses have been implemented on Kismet. It is derived from the evolutionary, cross-species, and social functions hypothesized by Plutchik [19].

Communicative expression. Each entry of this table has a goal-achieving behavioral component accompanied by a corresponding affective display. For some, the expressive display addresses both aspects of the emotive response when it serves a communicative function designed to elicit a desired response from a human. For instance, the robot exhibits sorrow upon the prolonged absence of a desired stimulus. This may occur if the robot has not been engaged with a toy for a long time. The sorrowful expression is intended to elicit attentive acts from the human analogous to how an infant's cries elicit nurturing responses from its caregiver. In a number of human robot interaction studies with Kismet [22], we have found this to be quite effective as people find pleasure in cheering up the robot. Kismet also uses its expressive displays to encourage people to slow down or back off a bit if they are crowding its cameras, moving too fast for the robot to perceive them, etc. This allows the robot to tune the human's behavior so that it is appropriate for it [23].

Decision making. Another class of affective responses relates to decision making for improved behavioral performance. For instance, a successfully accomplished goal elicits a joyful response that is reflected by a smile on the robot's face, whereas delayed progress results in a state of frustration that is reflected by a stern expression. As Kismet grows increasingly "frustrated" it lowers the activation level of the active behavior. This makes it more likely to switch to another behavior that could have a greater chance of success of achieving the current goal.

Goal prioritization. Emotion and affect also plays an important role in helping to prioritize goals and deciding when to switch among them. For instance, Kismet has several protective responses that allow it to quickly switch from engagement behaviors to avoidance behaviors once an interaction becomes too intense or turns potentially harmful. This fear response can "hijack" the behavior and motor systems to rapidly respond to the situation. Affective signals arising from the drives also bias which behaviors become active to satiate a particular motive. Hence, affective influences allow the robot's behavior to be flexible and opportunistic.

Biasing attention. Several of Kismet's emotive responses bias the robot's attention toward desired stimuli (e.g., those relevant to the current goal) and away from irrelevant stimuli. For instance, Kismet's exploratory responses include visual search for desired stimulus and/or maintaining visual engagement of a relevant stimulus. Kismet's visual attention system directs the robot's gaze to the most salient object in its field of view. The overall salience of a stimulus is a combination of its raw perceptual salience (e.g. size, motion, color) and its relevance to the current goal. Kismet's level of interest biases it to focus attention upon a relevant stimulus even when it has a lower raw visual salience over another more visually salient but less relevant stimulus. Alternatively, Kismet's disgust response allows it to reject and look away from an undesired stimulus. This directs the robots gaze to another point in the visual field where it might find a desired stimulus, and provides an expressive cue that tells the human the robot wants something else. We have found that people are quick to determine which stimulus the robot is after and readily present it to Kismet.

IX. Conclusion

From the robot's point of view, these examples illustrate how Kismet's emotion system works in concert with its cognitive system to address its competing goals and motives given its limited resources and faced with the inherent complexity and uncertainty of interacting with people in relatively unconstrained scenarios [10]. The emotion system achieves this by assessing and signaling the value of immediate events in order to appropriately regulate and bias the cognitive system to help focus attention, prioritize goals, and to pursue the current goal with an appropriate degree of persistence and opportunism. They protect the robot from intense interactions that may be potentially harmful.

In many cases, Kismet must work in partnership with the human to achieve its goals. To do so, it must communicate its motives and goals to the person in an effective way through expressive cues and goal-directed behavior. As a result, human and robot work together, mutually regulating the others behavior through social cues, to establish and maintain a suitable interaction where the robot's motives and goals are satisfied in a flexible and timely manner. This benefits the robot.

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

References

[1] C. Darwin, *The Expression of the Emotions in Man and Animals*, John Murray, London, UK, 1872.

[2] J. Gould, *Ethology*, Norton, 1982.

[3] D. Norman, A. Ortony, and D. Russell, "Affect and machine design: Lessons for the development of autonomous machines," *IBM Systems Journal*, vol. 42, no. 1, pp. 38–44, 2003.

[4] R. Picard, *Affective Computation*, MIT Press, Cambridge, MA, 1997.

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

[6] R. Picard, *Emotions in Humans and Artifacts*, chapter What does it mean for a computer to "have" emotions, MIT Press, 2001.

[7] D. Norman, "How might humans interact with robots," Keynote address to the DARPA/NSF Workshop on Human-Robot Interaction, San Luis Obispo, CA, September 2001.

[8] P. Ekman, "Are there basic emotions?," *Psychological Review*, vol. 99, no. 3, pp. 550–553, 1992.

[9] K. Scherer, "Evidence for both universality and cultural specificity of emotion elicitation," in *The Nature of Emotion*, P. Ekman and R. Davidson, Eds., pp. 172–175. Oxford University Press, New York, NY, 1994.

[10] C. Breazeal, *Designing Sociable Robots*, MIT Press, Cambridge, MA, 2002.

[11] Juan Velasquez, "When robots weep: A mechanism for emotional memories," in *Proceedings of th 1998 National Conference on Artificial Ingelligence, AAAI98*, July 1998, pp. 70–75.

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

[13] N. Tinbergen, *The Study of Instinct*, Oxford University Press, New York, NY, 1951.

[14] K. Lorenz, *Foundations of Ethology*, Springer-Verlag, New York, NY, 1973.

[15] C. Breazeal and L. Aryananda, "Recognition of affective communicative intent in robot-directed speech," *Autonomous Robots*, vol. 12, no. 1, pp. 83–104, 2002.

[16] A. Fernald, "Intonation and communicative intent in mother's speech to infants: Is the melody the message?," *Child Development*, vol. 60, pp. 1497–1510, 1989.

[17] N. Vlassis and A Likas, "A kurtosis-based dynamic approach to gaussian mixture modeling," *IEEE Transactions on Systems, Man, and Cybernetics: Part A*, vol. 29, 1999.

[18] C. Izard, "Four systems for emotion activation: Cognitive and noncognitive processes," *Psychological Review*, vol. 100, pp. 68–90, 1993.

[19] R. Plutchik, *The Emotions*, University Press of America, Lanham, MD, 1991.

[20] C. Breazeal, "Emotive qualities in lip synchronized robot speech," *Advanced Robotics*, vol. 17, no. 2, pp. 97–113, 2003.

[21] C. Smith and H. Scott, "A componential approach to the meaning of facial expressions," in *The Psychology of Facial Expression*, J. Russell and J. Fernandez-Dols, Eds., pp. 229–254. Cambridge University Press, Cambridge, UK, 1997.

[22] C. Breazeal, "Emotion and sociable humanoid robots," *International Journal of Human-Computer Studies*, vol. 59, pp. 119–155, 2003.

[23] C. Breazeal, "Regulation and entrainment for human-robot interaction," *International Journal of Experimental Robotics*, vol. 21, no. 10-11, pp. 883–902, 2002.

Cynthia Breazeal is an Assistant Professor of Media Arts and Sciences at the MIT Media Lab. She received her B.S. degree in electrical and computer engineering at University of California, Santa Barbara, and the M.S. and Sc.D. degrees in electrical engineering and computer science from the Massachusetts Institute of Technology (MIT), Cambridge, in 1993 and 2000 respectively. Her interests focus on human-like robots that can interact, cooperate, and learn in natural, social ways with humans.