Affect in Human-Robot Interaction

Ronald C. Arkin* and Lilia Moshkina**
*School of Interactive Computing, Georgia Tech, Atlanta, GA USA
**Naval Research Laboratory, Washington, DC USA

Abstract. More and more, robots are expected to interact with humans in a social, easily understandable manner, which presupposes effective use of robot affect. This chapter provides a brief overview of research advances into this important aspect of human-robot interaction.

Keywords: human-robot interaction, affective robotics, robot behavior

I. Introduction and Motivation

Humans possess an amazing capability of attributing life and affect to inanimate objects (Reeves and Nass 96, Melson et al 09). Robots take this to the next level, even beyond that of virtual characters due to their embodiment and situatedness. They offer the opportunity for people to bond with them by maintaining a physical presence in their world, in some ways comparable to other beings, such as fellow humans and pets.

This raises a broad range of questions in terms of the role of affect in human-robot interaction (HRI), which will be discussed in this article:

- What is the role of affect for a robot and in what ways can it add value and risk to human-robot relationships? Can robots be companions, friends, even intimates to people?
- Is it necessary for a robot to actually experience emotion in order to convey its internal state to a person? Is emotion important in enhancing HRI and if so when and where?
- What approaches, theories, representations, and experimental methods inform affective HRI research?

I.1 Roles of emotion in robotics

There exist at least two different roles for emotion in robotic systems. The first, which will only be briefly discussed, is to serve as an adaptive function that increases the probability of correct behavior, some of which may relate to survival of an agent (human or robotic) in its environment. The second is for the benefit of the human when interacting with a robot by providing a means and mechanism for increasing the bandwidth in communication using non-verbal methods to create a more effective and stronger relationship between artifact and person.

I.1.1 Adaptive behavior for survival

Moravec (1988) notes that humans may even perceive emotions in robots even without deliberately modeling them: for example, if a robot backs away from a staircase it might be interpreted as a fear of falling by a person observing. Braitenberg (1984), using a series of Gedanken (thought) experiments, also demonstrates that vehicles can exhibit love, fear, and aggression that is attributed to them solely by virtue of human observation. These perceived emotions are likely to attune the robot more closely to its environment and thus enhance its survivability, but are not geared expressly for human-robot interaction. People have a natural propensity to anthropomorphize artifacts (Reeves and Nass 1996) even if there was no deliberate intent by the designer to do so.

I.1.2 Human-robot interaction

Many other researchers, some of which are discussed in more detail below, have chosen to deliberately embed explicit models of affect into robots, with the express purpose of enhancing the relationship between the human and robot, and in some cases with the explicit goal of fostering a strong attachment by a person to the artifact. The underlying goal here is to produce a robotic platform that can be a friend or even a life-long companion to a human (Arkin et al

2003), and in some cases even approach the possibility of intimate human-robot relations (Levy 2008).

I.2 Definitions in Context

Definitions for nebulous affective terms such as emotions can be debated ad infinitum (Arkin 05). This volume undoubtedly addresses this in other chapters. We should note, however, we take a solipsist stance, i.e., that robots do not need to experience affective phenomena in the same way as humans do or even at all, in order for them to be perceived as possessing them. So no claim is made that the robot actually experiences emotions, but rather that the goal of affective human-robot interaction is to *convey* the perception to a person that it does. While this may be unsatisfying to a philosopher, it is a pragmatic solution to the roboticist, where affect lies in the eye of the beholder.

I.3 A few short exemplars

In order to carry out this illusion, many psychological models of human affect have been explored. Two examples that have had commercial success are described.

I.3.1 Aibo

Aibo (Fig. 1 Left) was a robotic dog marketed by Sony Corp. from 1999-2006 and sold hundreds of thousands of units. It was intended to serve as a long-term companion and friend. It incorporated an instinct-emotional model, which ranged from an initial simplistic variable-based version to considerably more complex forms in some of its variants (Arkin et al 2003). This model was later incorporated into Sony's QRIO humanoid robot as well.

Figure 1: (Left) AIBO (Right) Paro. (Photographs from Wikipedia Commons.)

I.3.2 Paro

Paro on the other hand is a therapeutic robot for the elderly, which is also commercially available (Figure 1 Right). It has been evaluated in terms of physiological, psychological, and social benefits to the human it interacts with (Shibata 12). Paro yields its benefits by eliciting emotional responses in its users through direct physical interaction, and appears not to rely on a sophisticated internal emotional model.

II. Affective Robotics

Research in affective HRI has come a long way since the early forays of Moravec (1988), Tolman (Endo and Arkin 2001), Grey Walter (Holland 2003), and others, both in terms of breadth and depth. The following subsections discuss a number of representative recent examples showcasing a wide variety of approaches to robot affect.

II.1 Affective Models and Architectures

Several roboticists take a systematic approach of incorporating affective models into robotic architectures. Two such systems are described below: the TAME framework, which stands for

Traits, Attitudes, Moods and Emotions (Moshkina et al 2011, Moshkina 2011) and the DIARC architecture, which stands for Distributed Integrated Affect Cognition and Reflection (Scheutz et al 2007). Although these two systems differ in a number of important aspects, they both emphasize including affect as an integral part of a robotic architecture, and have been designed with the goal of facilitating overall human-robot interaction.

II.1.1 TAME: Traits, Attitudes, Moods, Emotions

The TAME framework is comprised of four psychologically-inspired interrelated affective phenomena that provides mechanisms for affect generation, affective behavior modification, and affect expression. TAME is platform-independent, and can work with a variety of robot architectures, although it is particularly well-suited for the behavior-based paradigm (Arkin 1998). Given relevant perceptual input, such as the categories of visible objects and distances to them (stimuli and their strengths), as well as some internal state information (e.g., battery level) and environmental conditions (e.g., light and noise levels), the TAME system produces situation-appropriate affect intensities, which in turn modify currently active task behaviors through parameter-adjustment (Figure 2 presents a conceptual overview).

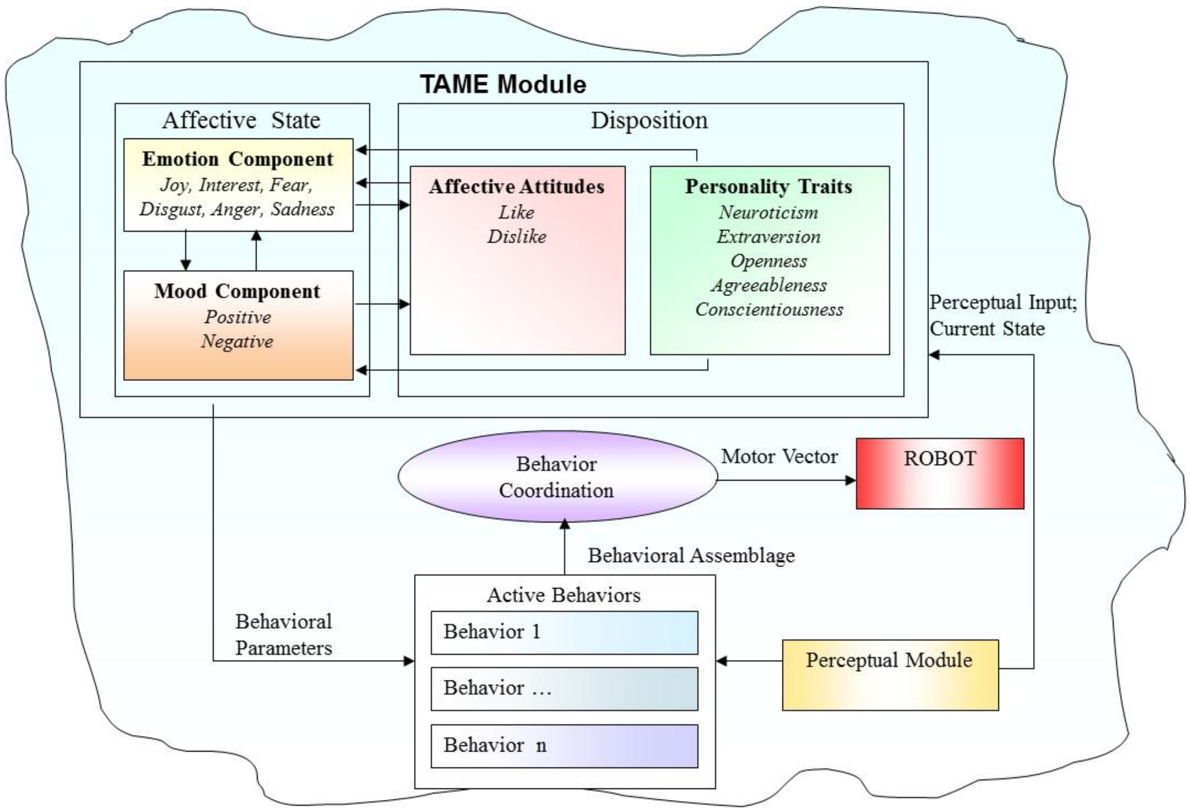

Figure 2: Conceptual view of the TAME framework (after Moshkina 2011).

The affective components comprising the framework provide a comprehensive, time-varying base for a robot, and differ with respect to duration (from almost instantaneous to life-long), and object-specificity (from very specific to diffuse and global). Emotions and moods constitute a robot's dynamically changing, transient affective state (object-specific and short-term for emotions, and diffuse and prolonged for moods). Moods provide an affective background, or "emotional color", and can vary cyclically, whereas emotions can be viewed as "phasic perturbations on this background activity" (Davidson 1994). In contrast, personality traits and attitudes are more or less time-invariant, and define general dispositions to behave and process information in certain ways. Similar to emotions, affective attitudes (sentiments) are object-specific; however, unlike emotions, they refer to ways of seeing and treating an object rather than to momentary responses, thus guiding behavior towards desirable goals and away from aversive objects. Finally, personality refers to enduring individual differences in behavior and information processing of a more general, object-independent kind, and serve as an adaptation mechanism to specialized tasks and environments.

The TAME framework was implemented as an independent software module, and integrated within *MissionLab*, a multiagent mission specification and execution robotic software toolset (MacKenzie et al 1997[1]), based on AuRA, a hybrid reactive-deliberative robotic architecture (Arkin and Balch 1997). Aspects of the resulting system were tested on Aldebaran Robotics' Nao biped humanoid platform in two human-robot experiments with over 70 participants. In one of these studies, the impact of Negative Mood and the emotion of Fear was assessed in a mock-up search-and-rescue scenario, where the participants found the robot expressing affect more compelling, sincere, convincing and "conscious" than its non-affective counterpart. Another study showed that different robotic personalities are better suited for different tasks: an extraverted robot was found to be more welcoming and fun for a task as a museum robot guide, where an engaging and gregarious demeanor was expected; whereas an introverted robot was rated as more appropriate for a problem solving task requiring concentration (Moshkina 2011).

II.1.2 DIARC Architecture

Unlike system-independent TAME, DIARC is an example of a novel robotic architecture incorporating general affect throughout its functional components. The architecture integrates cognitive capabilities (e.g., natural language understanding and action planning) and lower level activities (such as perceptual processing, feature detection and tracking, etc.) In this system, positive and negative affect (not differentiated into separate phenomena) plays a vital role in goal

[1] MissionLab is freely available for research and development and can be found at http://www.cc.gatech.edu/ai/robot-lab/research/MissionLab/

and task selection, by changing utilities of actions currently under consideration for selection, based on a short-term history of failures that produce increases in negative affect, and successes that produce increases in positive affect (Scheutz and Schermerhorn 2009). This mechanism serves as a kind of affective memory, allowing a robot to take into account past information without perfect knowledge of prior probabilities, a mechanism possibly used by humans in their decision-making.

In addition to affective goal action selection which takes place in every functional component of the architecture, DIARC also provides means for affect recognition, appraisal, and affective expression generation. DIARC has been used extensively as a testbed for research in HRI, including a number of human-subjects experiments and AAAI robot competitions. In one experiment (Scheutz, Schermerhorn and Kramer 2006), subjects and robots were paired in the context of a hypothetical space exploration scenario, where they had to work together to achieve a common goal. The results showed that when the robot and its human teammate were physically co-located, the robot's expression of anxiety over its lowering battery level (via voice) led to a performance advantage.

II.2 Socially Interactive Affective Robots

The two aforementioned affective robotic systems fall under the general umbrella of socially interactive robots, defined in (Fong, Nourbakhsh, and Dautenhahn 2003) as "robots for which social interaction plays a key role". A wide variety of other affective robotic systems also follow this research paradigm, focusing on affect as a key human social characteristic.

II.2.1 Robot Emotions

Emotion has been by far the most frequent affective phenomenon modeled in robots, though in some cases the distinction between emotion per se and other related phenomena has been blurred. The research in this category varies in terms of emotion generation mechanisms, modes of emotional expression, and underlying psychological and/or engineering approaches. Due to its practical applicability, facial emotional expressiveness has received a lot of attention, ranging from realistic robot heads to schematic faces to imaginary animals. Hanson Robotics android head "Einstein" (Wu et al 2009) is a good example of more or less realistic facial expressivity; the system is capable of learning and producing a large number of facial expressions based on Ekman's Facial Action Coding System, FACS (Ekman and Friesen 1978). Another system using FACS for displaying emotional expressions, though not as physically complex, is the expressive robotic head EDDIE (Sosnowski et al. 2006), capable of displaying affect based on the circumplex model of emotion (Posner, Russell and Peterson 2005). A highly stylized socially interactive robot head, ERWIN (Murray et al. 2009), expresses 5 basic emotions generated through modulation of hormonal-like parameters based on the context of current interactions. Farther down on the less realistic axis, an imaginary animal-like Huggable robot Probo (Goris et al. 2009) designed specifically for children, also produces emotional expressions based on the circumplex model of affect.

Although facial emotional display is the primary cue for emotion recognition in humans, not all physical platforms allow for this capability, lacking either facial motors, or heads altogether. A number of researchers addressed this challenge by designing for non-facial emotional expressivity. For example, Robovie-mini R2 and Robovie M (Nakagawa et al. 2009) are equipped with a method to control affective nuances by mapping dimensions of valence and arousal onto velocity and extensiveness of motion and body posture. In (Park et al. 2010), expressions of Fear and Joy, as well as Introversion and Extraversion, were achieved on a biped humanoid robot Nao (Aldebaran Robotics) through a combination of body posture and characteristic kinesics, that were successfully recognized in an online survey. For an extensive survey of non-facial non-verbal affective robot expressions the reader is directed to (Bethel and Murphy 2008).

II.2.2 Beyond Emotions: Multiple Affective Phenomena in Robots

A small subset of affective robotic systems differentiates between emotion and other affective phenomena, making the resulting affective capabilities richer and more compelling. Due to space limitations, apart from the aforementioned TAME, only four of these systems will be showcased here.

Roboceptionist – a combination of Emotions, Moods and Attitudes

This affective system (Kirby et al. 2010) was implemented on a virtual robot face placed on a rotating monitor at a receptionist's desk. It was used to interact with people on a daily basis for a prolonged time, and incorporated a generative model of affect consisting of emotions, moods, and attitudes. The affect is expressed through animated facial expressions and a priori composed narrative, rather than body language or mobility. The categorical emotions modeled in the system are joy, sadness, disgust, and anger, and are generated in response to interaction with people and displayed immediately after an eliciting event. The robot's moods are primarily caused by its personal history and "live" events. Values for moods are assigned to the storyline by dramatic writers and are influenced by the emotions the robot experienced during the day. Finally, attitudes are represented as a long-term mood associated with a particular person or thing, where each person who visits the robot may cause various emotional responses which, through mood modulation, influence the "opinion" of this person. In addition, familiarity with the person influences the robot's attitude towards that person.

A number of experiments have been conducted testing the components of this affect model. An on-line emotion recognition survey showed that people were able to detect differences between the robot's emotional expressions and differentiate between their intensities. Another study examined the influence of the robot's mood on people's interaction during a longer term (nine weeks, during which the robot was typically operating 8 hours per day, 5 days per week). During "low traffic" weeks, people interacted with the robot in positive mood for a shorter period of time than with the robot in neutral mood; in contrast, during "high traffic" weeks, where there were significantly more visitors, the robot in neutral mood elicited the least amount of interaction. This model is psychologically inspired to a certain extent, but relies heavily on input from the designers who write the robot's "life" story.

Waseda Eye No. 4 – a combination of Emotions, Moods and Personality

The latest incarnation of the robot, Waseda Eye No.4 Refined, combines emotions, moods, and personality (Miwa et al 2001, Miwa et al 2004). The overall goal of the system is to achieve smooth and effective communication for a humanoid robot. The Emotion space is defined along three dimensions: activation, pleasantness, and certainty. Emotions are represented as second order differential equations, based on laws of motion, and are influenced by three emotion coefficient matrices: Emotional Inertia, Emotional Viscosity, and Emotional Elasticity. The stimuli for emotion generation is extensive and includes visual (e.g., target is near), tactile (pushed, stroked, etc.), auditory (loud sound), temperature and olfactory (alcohol, smoke, etc.). The personality of the robot consists of Sensing and Expression Personalities. The Sensing Personality provides a mapping from sensory input to emotion generation as well as influences emotion duration and decay via the emotion coefficient matrices. The Expression Personality determines a particular emotional expression (Miwa et al 2001). Finally, mood is represented along pleasantness and activation axes. The mood pleasantness component is an integral of the emotion vector, and its activation component is based on an internal clock (Miwa et al 2004). The resulting emotional expression is not limited to the face, but also includes neck, waist, and arms; the speed of the motion is also varied depending on the emotion.

Although many elements of this system are not psychologically or biologically founded, it provides a few interesting mechanisms, such as modeling personality's influence on emotion via a variety of coefficient matrices and using internal-clock activation component in moods. No extensive human-robot interaction studies have been conducted to date to evaluate this system.

Combining Emotions and Moods on iCat

This research group (Leite et al 2008) implemented emotional reactions and moods on the Philips iCat robot within the context of a chess game. Emotional reactions were modeled as an "emotivector" – an anticipatory system that generates an affective signal resulting from a mismatch between the expected and sensed values of the sensor to which it is coupled. Mood is expressed as a less intense affective state, where positive values are associated with good scores in the game, and negative are related to bad scores. Moods are filtered over time, and are explicit when emotional reactions are not occurring. In two HRI experiments, it was found that emotional behavior of the robot helps users to have a better perception of the game (Leite et al 2008). Additionally, a later study (Castellano et al 2009) suggested that when the iCat displayed facial expressions during a game of chess, the level of user engagement towards the robot increased.

Combining Emotions and Motivational Drives in Kismet

The robotic creature Kismet (Breazeal 2003) is one of the earliest and most influential affective robotic systems. It is modeled after an infant and is capable of proto-social responses, providing an untrained user with a natural and intuitive means of communication. Kismet's motivation system consists of drives (motivations) and emotions, where emotions are a result of its affective state. The affective space is defined along three dimensions: arousal, valence and stance; each emotion is computed as a combination of contributions from drives, behaviors, and percepts. The motivation system plays a role in the behavior selection process and attention selection process, as well as providing activation for facial emotional expressions and speech.

II.3 Affect for Enhancing Robotic Behavior

The affective robotic systems described until now have all had a common goal for the inclusion of affect – i.e., to facilitate human-robot interaction. However, this review would be remiss if it ignored the efforts which did not have HRI as its primary focus; where the general projected improvements in performance due to the addition of affect may well prove useful in making potential interaction with humans more robust.

With a focus on improving the robot's behavior through decision-making, learning, or action selection, a number of researchers used the fuzzy logic approach to emotion generation. In El-Nasr et al. (2000)'s emotional system for decision-making in mobile robots, the emotional states have no definite boundaries, and are represented by fuzzy sets with intensities of low, medium, and high. Their values are generated according to fuzzy logic inference rules and the OCC model (Ortony, Clore and Collins 1988), and are based on goals and expectations where at different intensities the same emotion can trigger different actions. In another fuzzy logic-based system (Hashimoto, Hamada and Akazawa 2003), Fuzzy Cognitive Maps (FCM) are used to represent the generation and effects of emotional states. FCMs allow a robot to learn associations between stimuli and emotional states, as well as between emotions and tasks. Finally, (Yu and Xu 2004) present an emotional system consisting of four fuzzy emotions (Sad, Lonely, Disgust, and Fear) and four Sensory inputs (Energy, Friendship, Cleanness, and Brightness). Emotions are based on both sensor input and current emotional history and can influence behavior selection by increasing/decreasing corresponding action-selection weights.

Murphy and her colleagues (Murphy et al. 2002) describe the use of emotions to control a group of robots working on interdependent tasks by dynamically adapting current behaviors to the context or changing the set of active behaviors altogether. The emotion model is based on Scherer's multilevel process theory of emotions (Leventhal and Scherer 1987). The Emotional State Generator (a finite state machine) accepts measures of task progress as input, then emotion influences the task selection of the Behavioral State Generator. The following advantage of using emotions was noted - they help break cyclic dependency problems without centralized planning and minimum communication.

Another control architecture for autonomous mobile robots which incorporates biologically-inspired artificial emotions was implemented by (Lee-Johnson and Carnegie 2007). Five emotions modeled in the system are fear, sadness, anger, surprise, and happiness. They are characterized by certain elicitation/response patterns – e.g., fear is invoked if the robot is damaged, and anger if progress towards a goal is obstructed. Once elicited, emotions modulate a robot’s planning and control parameters providing bias towards certain drives without overtly controlling the behavior; e.g., anger helps achieve the current goal even at the expense of secondary considerations. This model was implemented on a simulated version of MARVIN, a custom-built mobile robot, and was shown to have certain advantages in a navigation task, such as fewer collisions, and greater exploration coverage.

III. Methods and Metrics - Measures of Success in Affective HRI

One of the major challenges facing affective HRI is effective testing and evaluation. In the task- or function-oriented areas of HRI (such as collaborative endeavors between people and robots, or learning by imitation) measuring robot performance is more or less straightforward. In the case of affective robots, however, often it is not the robot's performance per se that needs to be evaluated, but rather the social response the robot invokes in people it interacts with. These could be reflected in their subjective impressions (measured through self-assessments), behavioral responses and expressions (obtained through observation), certain physiological responses, or in objective difference in human task performance due to the presence of robot affect.

III.1 Self Assessments

These are subjective evaluations used to uncover people's perceptions of and attitudes towards their interactions with robots. In the HRI community, these methods of evaluation commonly include: Likert-style questionnaires designed for evaluating specific goals of a particular study (often applied in an ad hoc manner in this new field (Bartneck et al 2009)); reusable semantic differential scales or other psychometric scales for measuring certain concepts relevant to human-robot interaction (also designed specifically for use in HRI); and established psychological and sociological measures, borrowed from corresponding research communities.

III.1.1 Existing psychometric tests

These methods have the advantage of having been tested and validated on a large number of subjects; however, only a few of them have been tested to date with regards to affective robots. Two such measurement scales, particularly suitable for the affective HRI domain, are: (1) A brief version of Goldberg's Unipolar Big-Five Markers (provides personality trait assessment)

(Saucier 1994); and (2) Positive Affect/Negative Affect Schedule (measures current mood state) (Watson et al. 1998). These two instruments have been used successfully for assessing both subjects' and robot's personality and mood states in a number of HRI experiments conducted as part of TAME evaluations. The reader is referred to (Moshkina, 2011) for further details on their use and recommendations for their application in HRI.

III.1.2 HRI-specific tools

Social robotics is a very young field, and only a few reusable self-assessment tests are currently in existence. One of the most widely used ones is the Negative Attitudes towards Robots Scale (NARS), developed and tested by (Nomura and Kanda 2003, Nomura et al 2008). This scale measures general negative attitudes towards robots via three subscales: Situations and Interactions with Robots, Social Influence of Robots, and Emotions and Interaction with Robots, with each subscale item given as a Likert-style question.

Bartneck et al. (2009) present an overview of other existing scales which have been successfully used in HRI experiments and have acceptable internal reliability. These scales (most of them translated by the authors into semantic differential scales from Likert scales) measure the concepts of Anthropomorphism, Animacy, Likeability, Perceived Intelligence, and Perceived Safety of Robots.

Finally, two recent HRI studies (Moshkina 2011, Moshkina 2012) presented 7 alternative semantic differential scales for measuring concepts of relevance to affective HRI. These 5-item scales assess the following concepts with acceptable internal consistency reliability (Cronbach's Alpha for each 0.7 or higher): Persuasiveness, Naturalness, Understandability, Appropriateness, Welcome, Appeal, Unobtrusiveness and Ease. These scales were specifically designed with the goal of facilitating reuse and replicability in future HRI experimentation.

Although self-assessments are among the most commonly used methods of evaluation in HRI studies and allow querying people's perceptions of their interaction directly, they suffer from lack of objectivity. They can be notoriously unreliable, as they reflect a large amount of individual differences, and make replication of results and comparison between different studies rather difficult.

III.2 Behavioral and Psychophysiology Measures

Behavioral measures are observational, and refer to an analysis of participants' micro- and macro-behaviors and speech utterances during interaction. In this case, the human-robot interactions are recorded; the human behaviors to watch for are carefully selected and accurately described, and then are extracted from the video either automatically or by independent human coders (c.f. Jeff Cohn's chapter in this volume). Although not specifically within the affective domain, (Dautenhahn et al. 2002, Dautenhahn and Werry 2002) successfully used a combination of quantitative and qualitative behavioral techniques in HRI experiments. The quantitative approach was based on an analysis of micro-behaviors presented as well-identifiable, low-level, action-oriented categories; the examples of such categories include touch, eye gaze and eye contact, handling (picking up, pushing), approach, moving away. The qualitative approach was based on Conversation Analysis (Psathas 1995) that provides a systematic analysis of everyday and institutional talk-in interaction. Although these methods avoid some of the biases inherent in self-assessments, the differences in individual behavioral styles, possible interpretation bias, and high time workload can be cited as weaknesses in using these behavioral measures.

Another way to avoid participant subjectivity is to measure certain physiological responses (such as heart rate, skin conductance and temperature) before, during and after the interaction; such responses can be correlated with a subject's emotional state and arousal level (c.f. Jennifer Healey's chapter in this volume). The primary advantage of this method is that participants usually cannot manipulate the response of their autonomic nervous system; therefore the results obtained by this means are free from self-report bias. Perhaps the main disadvantage of this method is the limitation as to what they can measure; for example, they cannot distinguish anger from joy, but rather report the overall level of arousal. This method works well when, for example, the level of human anxiety needs to be determined (Kulic and Croft 2006, Rani et al. 2004), and has been used especially for affect recognition, but it would need to be supplemented by other measures to obtain cross-validation and additional information.

III.3 Task Performance

Finally, human task performance metrics provide a fair amount of objectivity, as they allow quantifying benefits a particular robot type, behavior, or algorithm might have. This is accomplished through such variables as accuracy, performance success, task completion time, error rate, resource usage and others, depending on a particular task and scenario. One example of employing a task performance metric to evaluate the effectiveness of robot affect is presented in (Scheutz et al. 2006). In their study, the authors measured changes in task performance as a result of a robot's expression of anxiety during an exploration scenario. In particular, as the robot's anxiety increased (expressed by changes in the robot's speech rate and pitch), the human participants were alerted to the impending deadline, and worked more efficiently.

Another HRI study examined the effect of robot expressions of Negative Affect (mood) and Fear on human subjects' compliance with a robot's request to evacuate a potentially dangerous area (Moshkina 2012). This study, set as a mock-up search-and-rescue scenario, showed that the participants responded to the request earlier and moved faster and further in response to the affective robot when compared to one without affect. Although task performance metrics provide objective and easily quantifiable results, their use in affective HRI is far from trivial. The biggest challenge lies in predicting which types of tasks would directly or indirectly benefit from affective robotic behaviors, and how the people would respond to a robot's expressed affective state.

Each type of evaluation has its associated pros and cons, and we join Bethel and Murphy (2010) in advocating the inclusion of more than a single method of evaluation to obtain comprehensive understanding and convergent validity in assessments of affective HRI.

IV. Ethical Questions and Future Directions

The ethical issues confronting roboticists relate to the questions: what if we succeed? What if we are able to create robotic artifacts that more effectively interact with humans than other humans do? Could we engineer out the difficulties in relationships that people often encounter with each other, through the use of affective models (among other techniques)?

Philosophers and Social Scientists (e.g., (Sparrow 02, Turkle 11)) have written about the potential dangers of highly interactive robots in human lives, broaching a broad range of issues, including:

- The introduction of a deliberate misapprehension of the world in the aged. Does there not exist a fundamental right to perceive the world as it really is? (i.e., the robots are not alive, even if they appear to be to a human observer).
- The abrogation of responsibility between humans by relegating the role of caregivers to robots and the resulting impact on those being cared for and the caregiver alike.
- The deterioration of the fabric of society and human relationships in general by potentially creating artifacts that are more appealing than fellow humans to interact with.

There are no obvious answers to these questions at this time, and considerable discussion and precaution is wise as robots move into ubiquity in our lives perhaps during our lifetimes.

Nonetheless the frontier remains to be explored. There are interesting models and a new understanding arising from deeper insights in neuroscience and the role that emotions play (Gazzaniga 2005). The role of mirror neurons (Arbib 2012) may show new ways to elicit emotional response in both human and robot alike. Understanding secondary emotions, such as those governing moral judgment may lead to robots that can even outperform humans in ethical respects (Arkin and Ulam 2005). The list goes on and on. It is an exciting time for the young field of human-robot interaction, and affect plays a central role in its progression.

V. References

1. Arbib, M. (2012). *How the Brain Got Language: The Mirror System Hypothesis*. Oxford University Press.
2. Arkin, R. (1998). *Behavior-based Robotics*. MIT Press.
3. Arkin, R., & Balch, T. (1997). AuRA: Principles and Practice in Review. *Journal of Experimental and Theoretical Artificial Intelligence*, 9(2), 175-189.
4. Arkin, R., Fujita, M., Takagi, T., & Hasegawa, R. (2003). An Ethological and Emotional Basis for Human-Robot Interaction. *Robotics and Autonomous Systems*, 42 (3-4).
5. Arkin, R.C. (2005). Moving Up the Food Chain: Motivation and Emotion in Behavior-based Robots. In J. Fellous & M. Arbib (Eds.), *Who Needs Emotions: The Brain Meets the Robot (pp. 245-271).* Oxford University Press.

6. Arkin, R.C. & Ulam, P. (2009). An Ethical Adaptor: Behavioral Modification Derived from Moral Emotions. *IEEE International Symposium on Computational Intelligence in Robotics and Automation*, Daejeon, KR.
7. Bartneck C., K.D., Croft E., & Zoghbi S. (2009). Measurement instruments for the anthropomorphism, animacy, likeability, perceived intelligence, and perceived safety of robots. *International Journal of Social Robotics*, 1, 71-81.
8. Bethel, C. L. & Murphy, R. R. (2008). Survey of Non-facial/Non-verbal Affective Expressions for Appearance-Constrained Robots. *IEEE Transactions on Systems, Man, and Cybernetics, Part C – Applications and Reviews*, 38(1), 83-92.
9. Bethel, C.L., & Murphy, R.R. (2010). Review of Human Studies Methods in HRI and Recommendations. *International Journal of Social Robotics*, 2, 347-359.
10. Braitenberg, V. (1984). *Vehicles: Experiments in Synthetic Psychology*, MIT Press.
11. Breazeal, C. (2003). Emotion and Sociable Humanoid Robots. *International Journal of Human-Computer Studies*, 59, 119-155.
12. Castellano, G., Leite, I., Pereira, A., & McOwan, P. (2009). It's All in the Game: Towards an Affect Sensitive and Context Aware Game Companion. *Affective Computing and Intelligent Interaction Workshop*, Amsterdam, Netherlands.
13. Dautenhahn, K., & Werry, I. (2002). A Quantitative Technique for Analysing Robot-Human Interactions. *International Conference on Intelligent Robots and Systems*, Lausanne, Switzerland.
14. Dautenhahn, K., Werry, I., Rae, J., Dickerson, P., Stribling, P., & Ogden, B. (2002). Robotic Playmates: Analysing Interactive Competencies of Children with Autism Playing with a

Mobile Robot. In K. Dautenhahn, Bond, A.H., Canamero, L., & Edmonds, B. (Eds.), *Socially Intelligent Agents (pp. 117-124)*, Kluwer Academic Publishers.

15. Davidson, R.J. (1994). On Emotion, Mood and Related Affective Constructs. In Ekman P. & Davidson, R.J. (Eds.)*, Nature of Emotion: Fundamental Questions* (pp. 51-56). Oxford University Press.
16. Ekman, P., & Friesen, W.V. (1978). *Facial Action Coding System*. Palo Alto: Consulting Psychologist Press.
17. El-Nasr, M.S., & Yen, J., Ioerger, T. (2000). FLAME - A Fuzzy Logic Adaptive Model of Emotions. *Journal of Autonomous Agents and Multi-agent Systems*, 3(3), 219-257.
18. Endo, Y. & Arkin, R.C. (2001). Implementing Tolman's Schematic Sowbug: Behavior-based Robotics in the 1930s. *IEEE International Conference on Robotics and Automation*, Seoul, Korea.
19. Fong, T., Nourbakhsh, I., & Dautenhahn, K. (2003). A Survey of Socially Interactive Robots. *Robotics and Autonomous Systems*, 42, 143-166.
20. Gazzaniga, Michael S. (2005). *The Ethical Brain*. New York: Dana Press.
21. Goris, K., Saldien, J., & Lefeber D. (2009). Probo, a testbed for Human Robot Interaction. *ACM/IEEE International Conference on Human-Robot Interaction*, San Diego, USA.
22. Hashimoto, T., Hamada, T., & Akazawa, T. (2003). Proposal of Emotion-driven Control Model for Robot-Assisted Activity. *IEEE International Symposium on Computational Intelligence in Robotics and Automation*, Kobe, Japan.
23. Holland, O. (2003). The First Biologically Inspired Robots. *Robotica,* 21(4), 351-363.
24. Kirby, R., Forlizzi, J., & Simmons, R. (2010). Affective Social Robots. *Robotics and Autonomous Systems,* 58, 322-332.

25. Kulic, D., & Croft, E. (2006). Physiological and subjective responses to articulated robot motion, *Robotica*, 15(1), 13-27.

26. Lee-Johnson, C.P., & Carnegie, D.A. (2007). Emotion-based Parameter Modulation for a Hierarchical Mobile Robot Planning and Control Architecture. *IEEE International Conference on Intelligent Robots and Systems*, San Diego, USA.

27. Leite, I., Pereira, A., Martinho, C., & Paiva, A. (2008). Are Emotional Robots More Fun to Play With? *IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN),* Munich, Germany.

28. Leventhal, H., & Scherer, K. (1987). The Relationship of Emotion to Cognition: A Functional Approach to a Semantic Controversy. *Cognition and Emotion*, 1, 3-28.

29. Levy, D. (2008). *Love and Sex with Robots: The Evolution of Human-Robot Relationships*. Harper Perennial.

30. MacKenzie, D.C., Arkin, R.C., & Cameron, J.M. (1997). Multiagent Mission Specification and Execution. *Autonomous Robots*, 4(1), 29-52.

31. *Manual for MissionLab Version 7.0*. Available: http://www.cc.gatech.edu/ai/robot-lab/research/MissionLab/. Georgia Tech Mobile Robot Laboratory, 2007.

32. Melson, G. F., Kahn, P. H., Jr., Beck A., & Friedman, B. (2009). Robotic Pets in Human Lives: Implications for the Human-Animal Bond and for Human Relationships with Personified Technologies. *Journal of Social Issues*, *65*, 545-567.

33. Miwa, H., Takanishi, A., & Takanobu, H. (2001). Experimental Study on Robot Personality for Humanoid Head Robot. *IEEE International Conference on Intelligent Robots and Systems*, Maui, USA.

34. Miwa, H., Etoh, K., Matsumoto, M., Zecca, M. Takanobu, H., Rocella, S., Carrozza, M., Dario, P., & Takanishi, A. (2004). Effective Emotional Expressions with Emotion Expression Humanoid Robot WE-4RII. *IEEE International Conference on Intelligent Robots and Systems*, Edmonton, Canada.
35. Moravec, H. (1988). *Mind Children: The Future of Robot and Human Intelligence*, Harvard University Press.
36. Moshkina, L.; Park, S.; Arkin, R.C.; Lee, J.; & Jung, H. (2011). TAME: Time Varying Affective Response for Humanoid Robots. *International Journal of Social Robotics,* 3(3), 207-221.
37. Moshkina, L. (2011). An integrative framework of time-varying affective robotic behavior. *Ph.D. Dissertation*, Georgia Institute of Technology.
38. Moshkina, L. (2012). Improving Request Compliance through Robot Affect. *AAAI Conference on Artificial Intelligence*, Toronto, Canada.
39. Murphy, R.M., Lisetti, C., Tardif, R., Irish, L., & Gage, A. (2002). Emotion-based control of cooperating heterogeneous mobile robots. *IEEE Transactions on Robotics and Automation*, 18(5): p. 744-757.
40. Murray, J.C., Canamero, L. & Hiolle, A. (2009). Towards a Model of Emotion Expression in an Interactive Robot Head. *IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*, Toyama, Japan.
41. Nakagawa, K.S., Ishiguro, H., Akimoto, T. & Hagita, N. (2009). Motion Modification Method to Control Affective Nuances for Robots. *IEEE International Conference on Intelligent Robots and Systems,* St. Louis, USA.

42. Nomura, T. & Kanda, T. (2003). On Proposing the Concept of Robot Anxiety and Considering Measurement of it. *IEEE International Workshop on Robot and Human Interactive Communication (RO-MAN)*, Millbrae, USA.

43. Nomura, T., Kanda, T., Suzuki, T., & Kato, K. (2008). Prediction of Human Behavior in Human-Robot Interaction Using Psychological Scales for Anxiety and Negative Attitudes towards Robots. *IEEE Transactions on Robotics*, 24(2), 442-451.

44. Ortony, A., Clore, G.L., & Collins, A. (1988). *The Cognitive Structure of Emotions*. Cambridge: Cambridge University Press.

45. Park, S., Moshkina, L., & Arkin, R.C. (2010). Recognizing Nonverbal Affective Behavior in Humanoid Robots. *IEEE International Conference on Intelligent Autonomous Systems*, Ottawa, Canada.

46. Posner, J., Russell, J., & Peterson, B. (2005). The Circumplex Model of Affect: An Integrative Approach to Affective Neuroscience, Cognitive Development, and Psychopathology. *Development and Psychopathology*, 17(3), 715–734.

47. Psathas, G. (1995). *Conversation Analysis*. Thousand Oaks: Sage.

48. Rani, P., Sarkar, N., Smith, C.A., & Kirby, L.D. (2004). Anxiety Detecting Robotic System—towards Implicit Human-Robot Collaboration. *Robotica*, 22(1), 85-95.

49. Reeves, B. & Nass, C. (1996). *The Media Equation: How People Treat Computers, Television, and New Media Like Real People and Places*. Cambridge University Press.

50. Saucier, G. (1994). Mini-Markers: a Brief Version of Goldberg's Unipolar Big-Five Markers. *Journal of Personality Assessment*, 63(3), 506-516.

51. Scheutz, M., Schermerhorn, P., & Kramer, J. (2006). The Utility of Affect Expression in Natural Language Interactions in Joint Human-Robot Tasks. *ACM International Conference on Human-Robot Interaction*, Utah, USA.
52. Scheutz, M., Schermerhorn, P., Kramer, J., & Anderson, D. (2007). First Steps toward natural human-like HRI. *Autonomous Robots*, 22(4), 411–423.
53. Scheutz, M., & Schermerhorn, P. (2009). Affective Goal and Task Selection for Social Robots. In Vallverdu, J. & Casacuberta, D. (Eds.), *Handbook of Research on Synthetic Emotions and Sociable Robotics: New Applications in Affective Computing and Artificial Intelligence* (pp. 74-87), IGI Global.
54. Shibata, T., Life Innovation with Therapeutic Robot, Paro. http://www.bbvaopenmind.com/article/en/60/life-innovation-with-therapeutic-robot-paro/, accessed August 2012.
55. Sosnowski, S., Bittermann, A., Kuhnlenz, K., & Buss, M. (2006). Design and Evaluation of Emotion-Display EDDIE. *IEEE International Conference on Intelligent Robots and Systems,* Beijing, China.
56. Sparrow, R. (2002). The March of the Robot Dogs. *Ethics and Information Technology*, 4(4), 305-318.
57. Turkle, S. (2011). *Alone Together: Why We Expect More from Technology and Less from Each Other*. Basic Books.
58. Watson, D., Clark, L.A., & Tellegen, A. (1998). Development and Validation of Brief Measures of Positive and Negative Affect: The PANAS Scales. *Journal of Personality and Social Psychology*, 56(6), 1063-1070.

59. Wu, T., Butko, N.J., Ruvulo, P., Bartlett, M.S., & Movellan, J.R. (2009). Learning to Make Facial Expressions. *IEEE International Conference on Development and Learning*, Shanghai, China.

60. Yu, C., & Xu, L. (2004). An Emotion-Based Approach to Decision-Making and Self Learning in Autonomous Robot Control. *5th World Congress on Intelligent Control and Automation,* Hangzhou, China.