AN INTEGRATIVE FRAMEWORK OF TIME-VARYING AFFECTIVE ROBOTIC BEHAVIOR

A Dissertation
Presented to
The Academic Faculty

By

Lilia V. Moshkina

In Partial Fulfillment
Of the Requirements for the Degree
Doctor of Philosophy in Computer Science

Georgia Institute of Technology
May, 2011

AN INTEGRATIVE FRAMEWORK OF TIME-VARYING AFFECTIVE ROBOTIC BEHAVIOR

Approved by:

Dr. Ronald C. Arkin, Advisor
College of Computing
Georgia Institute of Technology

Dr. Tucker Balch
College of Computing
Georgia Institute of Technology

Dr. Melody M. Jackson
College of Computing
Georgia Institute of Technology

Dr. Arthur D. Fisk
School of Psychology
Georgia Institute of Technology

Dr. Ayanna M. Howard
School of Electrical and Computer Engineering
College of Computing, Adjunct
Georgia Institute of Technology

Date Approved: March 31, 2011

ACKNOWLEDGEMENTS

This dissertation is a product of many years of work, and would not be possible without those who lent me their support, knowledge, and inspiration. First of all, my heartfelt gratitude goes to my husband, friend and colleague, Eric Martinson, for his unfailing encouragement, insightful ideas, hands-on help, and superior editing, not to mention taking excellent care of our young daughter Maya during the times I had to be away in order to conduct my experiments.

My academic advisor, Professor Ron Arkin, has been instrumental in bringing this research to completion, and I am very grateful for his direction, support, both intellectual and financial, and his almost infinite patience. I would also like to thank my dissertation committee for their guidance in a number of related domains and their invaluable comments. My special thanks go to Professor Melody Jackson, who has been a source of encouragement for many years, and who inspired me to enter into the Ph.D. program at GA Tech in the first place.

I would also like to thank Samsung Electronics Corporation for providing a two-year grant for implementation and testing of the *TAME* framework, as well as Hyunryong Jung and Chien-Ming Huang who were supported by this grant. I am especially grateful to Sunghyun Park, whose unflagging help and support went above and beyond duty and have been indispensable in completion of this dissertation.

Finally, I would like to express my deep gratitude to my friends and lab mates. My dear friend Natalia Dyomina has always been there for me, whether I needed encouragement, fresh ideas or statistics help, and my early conversations with her provided the inspiration for the theme of my research. My friends Ernest Merrill, Mario Romero and Sergey Demin were willing participants in all my experiments, and gave much appreciated moral support and excellent advice. Last but not least, my lab mates from Mobile Robot Lab: Zsolt Kira, Alex Stoytchev, Yoichiro Endo, Will Wagstaff and Brian Lee could be counted on for deciphering software and hardware intricacies, and making the lab a livelier place.

TABLE OF CONTENTS

List of Tables

List of Figures

SUMMARY

As robots become more and more prevalent in our everyday life, making sure that our interactions with them are natural and satisfactory is of paramount importance. Given the propensity of humans to treat machines as social actors, and the integral role affect plays in human life, providing robots with affective responses is a step towards making our interaction with them more intuitive. To the end of promoting more natural, satisfying and effective human-robot interaction and enhancing robotic behavior in general, an integrative framework of time-varying affective robotic behavior was designed and implemented on a humanoid robot. This psychologically inspired framework (TAME) encompasses 4 different yet interrelated affective phenomena: personality Traits, affective Attitudes, Moods and Emotions. Traits determine consistent patterns of behavior across situations and environments and are generally time-invariant; attitudes are long-lasting and reflect likes or dislikes towards particular objects, persons, or situations; moods are subtle and relatively short in duration, biasing behavior according to favorable or unfavorable conditions; and emotions provide a fast yet short-lived response to environmental contingencies. The software architecture incorporating the TAME framework was designed as a stand-alone process to promote platform-independence and applicability to other domains.

In this dissertation, the effectiveness of affective robotic behavior was explored and evaluated in a number of human-robot interaction studies with over 100 participants. In one of these studies, the impact of Negative Mood and 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. To conclude, multi-faceted robotic affect can have far-reaching practical benefits for human-robot interaction, from making people feel more welcome where gregariousness is expected to making unobtrusive partners for problem solving tasks to saving people's lives in dangerous situations.

1 INTRODUCTION

We are born, we grow, we develop our personality that influences our career choices, the choice of people we associate with, that shapes our life. We learn to love and hate, admire and despise, and we behave accordingly towards those individuals or things that invoked those attitudes. As our life goes on, we may fear spiders – and avoid them, rejoice in our accomplishments – and strive to do even better. Our moods change as the days go by: we may feel low when it's dreary outside, and creative and energized when the sun comes back out. Although we may not want to admit it, our lives are heavily influenced by our affective space: emotions, moods, attitudes, and personality.

Robots are not like us: they are not born, they don't grow, they don't need to make career choices or have accomplishments. Therefore, they don't need to be governed by affect or display affective behavior – or do they? If they are in danger, do they not need to avoid it lest they be destroyed or otherwise be rendered unusable? While performing a task, do they not need specific characteristics more suitable to this particular task than others? And as the robots become more and more prevalent in our lives, would they not need to interact with us in terms we know the best, social? By modeling a robot's affective space we could achieve a richness and effectiveness of robotic behavior that would be hard to accomplish otherwise.

What is an affective space comprised of? By definition, "affective" means "concerned with, arising from, relating to or influencing emotion" [1, 2]. There are a number of phenomena that can be classified as affective and which we include in affective space: personality, affective attitudes, moods and emotions themselves. Personality influences both generation and display of emotion (e.g., a fearful person tends to experience fear more often, and a happy person joy). Affective attitudes, or sentiments, refer to propensities to respond emotionally to specific persons, objects, or events [3]. Moods are diffuse, global, low-intensity affective states [4], and, finally, emotions are at the center of the definition of the term “affective”. Each of these phenomena performs a variety of distinct functions, occupies a separate place on the time continuum (e.g., personality is relatively time-invariant, whereas emotions are very brief in duration), and is highly interrelated with one another. Together, these phenomena have enabled biological systems to survive, adapt and develop through extensive interaction. Robotic systems display similar needs, and thus could benefit from similar affective mechanisms.

Our purpose in including a rich model of affect into robotic systems is then two-fold. First, it would serve to enhance the effectiveness of robotic behavior, contributing to a robot’s survivability and adaptability. The role of affect for survival has been long established in both animals and people. For example, personality is considered to provide a goodness-of-fit mechanism from the evolutionary standpoint [5, 6], and it plays a role in action selection and determines consistent patterns of behavior across situations [7]. Thus, modeling personality could facilitate adjustment to a variety of environments and matching heterogeneously behaving robots to tasks. Other affective phenomena play similarly important roles in our behavior [4, 8-11].

Secondly, as we expect robots to become a part of our everyday lives, modeling affect in robotic systems will provide for more natural, effective and satisfying human-robot interaction. Humans are inherently social creatures, and apply social rules not only to their interactions with one another, but also to those with non-human animals, and even inanimate objects. This propensity of people to anthropomorphize certain objects has been well established by Nass and his colleagues in an extensive set of experiments [12], which showed that people treat computers as social actors, whether they recognize it or not, and that even minimal cues evoke social responses. As affect plays a vital role in our social interactions (Oatley and Jenkins [13] call emotions “the language of human social life”), it would be beneficial for robots to be able to “speak” this language. Giving robotic systems the ability to communicate with humans affectively would not only allow people to extend their known social models to robots, but would also help robots invoke desired responses from their human interaction partners or even passers-by (e.g., assisting a stuck robot, or evacuating from a dangerous zone).

To the end of both enriching effectiveness of behavior-based robotic systems and promoting a more natural, effective and satisfying human-robot interaction, we propose an integrative framework of time-varying affective robotic behavior, *TAME*, in which the aforementioned phenomena are modeled as separate components within affective space, with explicitly defined interactions between them. *TAME* stands for *Traits* (personality), *Attitudes, Moods* and *Emotions*, the four components responsible for producing affective behavior. Although the framework itself is designed in support of both goals, the emphasis in this dissertation will be placed on the latter, namely, enhancing the interaction between humans and robots.

Given the predicted pervasiveness of robots in our daily lives and our social human nature, a number of researchers [14, 15] support the paradigm of robots as collaborative partners/companions and acknowledge the role of affect in such collaborative interaction between humans and robots. However, to the best of our knowledge, no current research initiatives intend to combine this wealth of affective phenomena into a single integrative time-varying framework. The research in this dissertation is also in line with the Robotics 2.0 initiative [16], which places the human in the loop. Robotics 2.0 proposes to expand the roles of robots to those of co-workers, co-inhabitants, and co-protectors, and applications of the *TAME* framework to this *Co-X* paradigm will be considered throughout the dissertation.

1.1 TERMINOLOGY

The following terms will be used throughout the dissertation:

- *Affective:* perceived in positive or negative terms; refers to processes or phenomena in humans and animals traditionally thought of as "emotional" or those that cannot be explained strictly in terms of cognition.
- *Affective Attitude*: a learned predisposition to respond in a consistently favorable or unfavorable manner with respect to a given object; a general and enduring positive or negative feeling about some person, object or issue.
- *Behavior-based*: consisting of individual behaviors and coordination mechanisms between them.
- *Emotion:* organized reaction to an event that is relevant to the needs, goals, or survival of the organism; is short in duration and noncyclical; is characterized by a high activation state and significant energy and bodily resources expenditure.
- *Mood:* a continuous, low-activation variable affective state, or "stream of affect"; is more inclusive than emotion, non-specific, longer in duration and can be cyclical.
- *Personality, traits:* consistent patterns in behavior over situations and time; developed as a goodness-of-fit mechanism from an evolutionary standpoint.

1.2 RESEARCH QUESTION

Does integration of coordinated time-varying affective processes (namely, emotions, moods, affective attitudes and personality traits) into behavior-based robotic systems generate more effective robotic behavior from the human-robot interaction standpoint?

1.2.1 SUBSIDIARY QUESTIONS

The subsidiary questions to help explore the main research question are as follows:

- *How can the aforementioned phenomena be modeled computationally in a robotic system, both individually and relative to each other?*

What are the psychological foundations for each of the components and their interactions? How can these phenomena be represented, generated, and applied to robotic behavior? What are their functions, and what is their relevance for robotic behavior? What are the interactions between them that can provide additional benefit beyond that of each individual component? Is there a need to include all four components into the framework? If not, then what components are desirable?

- *What are the implications for Human-Robot Interaction? Does complex affective robotic behavior lead to more natural, effective, and satisfying interaction between humans and robots?*

Does coordinated affect in robotic behavior improve the ease and pleasantness of a human interacting with a robot in more or less everyday activities? Would familiar affective responses make it easier for users (especially novices) to accept and enjoy the interaction? Would the complex affective behavior help maintain human's interest in interacting with robots, or even form attachments? Would it make the robots appear more natural, persuasive, comprehensible, and welcoming?

- *What are the metrics for evaluating affective robotic behavior?*

How can we evaluate the effectiveness of adding a rich model of affective behavior as it relates to human-robot interaction? Are there specific quantitative and qualitative metrics to accurately perform such an evaluation?

1.3 EXPECTED CONTRIBUTIONS

Expected contributions stem directly from the aforementioned research issues:

- A framework to augment behavior-based robotic systems with a variety of time-varying affective processes, namely emotions, moods, attitudes and personality;

- Means for affective communication to afford more natural human-robot interaction;
- Metrics for evaluating effectiveness of affective robotic behavior.

1.4 DISSERTATION OVERVIEW

The remainder of this dissertation is structured as follows. In chapter 2 we will present a review of relevant research in the areas of robotics and software agents. Chapter 3 will focus on psychological and mathematical foundations for each of the *TAME* components and the interactions between them. Chapter 4 will present a longitudinal human-robot interaction study conducted with the purpose of informing the design process. The architectural design and implementation of the *TAME* framework, as well as an online survey assessing recognition of nonverbal affective robotic behaviors, will be provided in Chapter 5. The first subsidiary question will be addressed in chapters 3 and 5. Two human-robot interaction studies designed to determine the usefulness of selected framework components will be discussed in detail in Chapter 6, which will address the second subsidiary question. Novel metrics for evaluating the effectiveness of affective robotic behavior, derived from the studies, will be presented in Chapter 7, where the third subsidiary question will be addressed. Finally, Chapter 8 will present conclusions and contributions.

2 RELATED WORK

Within the past two decades, influences from at least three separate fields: neuroscience, HCI and psychology, brought increased interest to the problem of affect in robots and computational agents. Neurology has established the active role of emotions in human cognition and behavior, e.g., LeDoux proposed that emotions are powerful motivators of behavior [17], and Damasio implicated emotions and feelings in such cognitive processes as reasoning and decision-making [9]. From HCI came a counterintuitive at first, but currently well-established proposition that humans treat computers as social actors [12]. In particular, people are polite to computers, are responsive (positively) to flattery and (negatively) to criticism, easily pick up various personality cues, even minimal, and accept computers as teammates [18]. Finally, in psychology, Ortony, Clore and Collins presented a Theory of the Cognitive Structure of Emotions that turned out to be highly amenable to computer implementation, and has been widely used to generate emotions in artifacts since 1988 [19]. This theory, often referred to as the OCC model, classifies emotion according to cognitive eliciting conditions and includes a rule-based system for generation of emotion types.

In her highly influential book “Affective Computing” Rozalind Picard advocates the usefulness of affect in computers (a term used quite broadly, to include computational agents, robots, wearables, etc.) [20]. She distinguishes between recognizing, expressing, and synthesizing affect. Although a wide variety of affect-recognition systems exists in both the robotics and computational agents communities (e.g., affective user modeling for entertainment, education and other domains [21-23], tangible interfaces [24, 25], and sensing and responding to user frustration and anxiety in computer interfaces [26] and robots [27]), this work won’t be covered here in detail as the main focus of this research is on synthesis and expression (including behavioral) in robotic systems.

To date, there has not yet been any systematic work on computational generation and application of all aspects of affect. In general, researchers have been trying out a wide variety of ideas for a number of purposes, ranging from improving dialogue believability in embodied conversational agents, overall believability in virtual characters and synthetic actors, to facilitating human-robot interaction and providing therapeutic psychological effects for humans, to enhancing robot learning and survival capabilities. The classification task of existing affective systems is further exacerbated by the fact that there is no consensus among psychologists and cognitive scientists as to what exactly affect is comprised of and what are its functions, generation and expression processes [3, 28-31]. This uncertainty resulted in a variety of approaches to computational models of affect, among which are biologically, ethologically, cognitively and neurologically inspired systems, in addition to those guided by primarily design and engineering considerations. However, for the most part, the inspiration is taken from a multitude of domains; therefore, in this chapter we provide a loose categorization of affect-related systems according to their domain and purpose, accompanied by their relevance to the research described in this dissertation. As our framework is mainly concerned with incorporation of four distinct but interrelated affective phenomena, namely, *personality Traits, affective Attitudes, Moods and Emotions*, the emphasis will be placed on those systems/models that combine the aforementioned phenomena. A brief overview of user studies assessing people's attitudes to and perceptions of social robots is also presented, along with various evaluation methods that resulted from such studies.

2.1 SOCIALLY INTERACTIVE AFFECTIVE ROBOTS

> **Reporter:**
> One gets the sense that he (HAL) is capable of emotional responses. When I asked him about his abilities I sensed a sort of pride…
> **Bowman:**
> Well, he acts like he has genuine emotions. Of course, he's programmed that way to make it easier for us to talk with him. But whether or not he has real feelings is something I don't think anyone can truly answer.
>
> *Movie, 2001 Space Odyssey*
> *Directed by Stanley Kubrick (1968)*

Socially interactive robots are defined in Fong et al. [15] as "robots for which social interaction plays a key role". In particular, it implies exhibiting certain "human social" characteristics, of which affect is an integral part: establishing/maintaining social relationships, expressing and/or recognizing emotion, and exhibiting distinctive personality and character, among others. The related research presented in this section focuses on affective capabilities of such socially interactive robots designed for improving human-robot interaction. Systems most relevant to this research (those combining multiple affective components) will be described first, followed by those combining emotions and more general motivations, finally followed by a variety of robotic systems in which only a single phenomenon (most often emotions) is modeled.

2.1.1 COMBINING EMOTIONS, PERSONALITY, MOODS OR ATTITUDES

The systems presented below incorporate, at least to a certain extent, a combination of affective phenomena.

2.1.1.1 Roboceptionist (Emotions, Moods and Attitudes)

Kirby et al. [32] present an affective model for social robots, which incorporates emotions, moods and attitudes. The model was implemented on a virtual robot face placed on a rotating monitor, and the affect was 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 are 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 emotions the robot experienced during the day. Finally, attitudes are represented as a long-term mood associated with a 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 person's attitude.

A number of experiments have been conducted to test 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 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.

2.1.1.2 Waseda Eye No. 4 Refined (Emotions, Moods and Personality)

This system was created by Miwa, Takanishi and Takanobu [33-35]. The latest incarnation of their robot, Waseda Eye No.4 Refined, combines emotions, moods, and personality. 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 (target is near, etc.), 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 [33].

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 [34]. 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.

2.1.1.3 Emotional Robot "Cherry" (Personality, Emotions and Moods)

Lisetti [36] proposes incorporating personality, emotions and moods into robotic systems with the goal of achieving "social expertise"; however, moods have not been incorporated as yet, and later work places the emphasis primarily on emotions [37]. To embed emotions, a multilevel theory [38] is used, in which the sensorimotor level is automatically activated, schematic level integrates sensorimotor processes with scripts of emotional situations, and the conceptual level is deliberative. In the hybrid reactive/deliberative robotic architecture (based on [39]), the sensorimotor level corresponds to the reactive layer, and effects parameters of actively running behaviors, whereas the schematic level, where emotions are represented as scripts, influences the set of active behaviors at the assemblage level.

In the implementation of the ActiveMedia Peoplebots office service robot Cherry, the emotions were expressed via facial expressions and speech. Their generation was achieved via an FSA, where the transitions between emotions occur as emotional inputs are received. This implementation suggests that emotions are ever present, with or without external stimuli. The only application of personality presented refers to the robot's experiencing various tendencies towards specific sets of emotions based on the preprogrammed personality traits.

As far as the emotions are concerned, this system is psychologically inspired; however, the inclusion of personality seems to be only incidental and does not affect behavior per se.

2.1.1.4 Character Robot Face (Emotions and Moods)

Fukuda et al. [40] also include the notions of emotions and moods in their Character Robot Face. Emotions are represented as semantic networks, and the combination of currently active emotions is deemed as mood. The currently implemented happy, neutral, and sad moods are in this case a summation of the experienced emotions of happiness or sadness (based on tactile – being petted or hit and visually – blue or orange objects perceptual stimuli). An interesting component in the system is a mood and task coordination mechanism: associating a human-given task with the current emotion, which biases the robot to pick the same task the next time the emotion is experienced. For example, once learned, the robot will "feel" happy while performing the task even without interacting with people (the internal trigger of happiness). In addition to their influence on action selection, another effect of emotions are expressions exhibited via a nonhuman face, modified from emotional expressions defined Ekman's Facial Action Coding System [41]. This system seems mainly engineering-driven and is somewhat limited in both the set of emotions present and the way the moods are generated.

2.1.1.5 iCat (Emotions and Moods)

Leite et al. [42] 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 that generates an affective signal resulting from the mismatch between the expected and sensed values of the sensor to which it is coupled to. 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 preliminary HRI experiments, it was found that emotional behavior of the robot helps users to have a better perception of the game [42]. Additionally, a later study [43] suggested that when the iCat displayed facial expressions during a game of chess, the level of user engagement towards the robot increased.

2.1.1.6 "Mental Commit" Robot Paro (Emotions and Moods)

Yet another robotic system that combines affective phenomena of emotions and moods is that of Wada et al. [44]. They propose "mental commit" robots that are designed to engender mental effects in humans, such as pleasure and relaxation, in their role of personal robots. Such artifacts can be used for Robot-Assisted Therapy. Their seal robot Paro, in the behavior-planning layer, has internal states that are labeled as emotions. They decay in time, are driven by interaction, and change behavioral parameters in the behavior-generation layer. The robot also has a diurnal rhythm with several functions that depend on it, e.g., sleep – such cyclic functions can be classified as moods. This influence of mood-like states on other functions separates this work from other research in the area, where moods are primarily used in emotion generation. A number of human user studies involving this robot have been conducted; however, the emphasis was on the robot as a whole, not on the affective mechanisms in particular [45, 46].

The systems reviewed in this subsection are summarized in **Table 1**:

Table 1: Robotic Systems with Multiple Affective Phenomena

System	Foundation	Emotions	Moods	Personality	Attitudes
Roboceptionist	Psychological/drama	Yes	Yes	No	Yes
Waseda Eye No 4 Refined	Design/ biological	Yes	Yes	Yes (only in regards to emotions)	No
Cherry	Psychological	Yes	No	Yes (only in regards to emotions)	No
Character Robot Face	Design/ biological	Yes	Yes	No	No
iCat	Design/psychological	Yes	Yes	No	No
Paro	Design	Yes	Yes	No	No

2.1.2 COMBINING EMOTIONS AND DRIVES

A number of researchers include emotions as part of a wider motivational system. One such system is Breazeal's robotic creature Kismet [47]. It can probably be considered the first socially interactive robot. Kismet is modeled after an infant and is capable of proto-social responses, providing an untrained user with 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.

In Arkin et al [48], a similar dimensional emotional model (in which the dimensions of arousal, valence and confidence are modeled) serves as part of an extensive ethology-inspired motivational system for the Sony robotic dog AIBO. The system utilizes homeostatic regulation mechanism: the internal variables (e.g., emotions) must be regulated and maintained within proper ranges. Emotions and drives are directly involved in action selection. Furthermore, a mechanism for "emotionally grounded symbols" is proposed in the Emotionally GrOunded Architecture (EGO) that allows learned associations to be formed between emotional experiences and grounded physical symbols.

Finally, in the robot MEXI (Machine with Emotionally eXtended Intelligence) [49], the Emotion Engine is composed of a set of basic emotions (positive that it strives to achieve, and negative that it strives to avoid) and homeostatic drives. Generated emotions influence the Behavior Engine, and actions respectively, only if they reach a certain threshold.

The systems described in this subsection include only one affective phenomenon – emotions. What makes these systems interesting for this research is that in addition to purely expressive role of emotions, they also examine their influence on wider range of behaviors.

2.1.3 OTHER SYSTEMS

Other robotic systems including a single dimension of affect (primarily emotions) in order to improve human-robot interaction include:

- The LEGO robot Feelix [50] capable of expressing a subset of basic emotions elicited through tactile stimulation;

- A social robot Sparky [51], in which different emotional states are expressed in a caricature robotic face and ambient motion, and in which emotional expressions are triggered by the operator;
- A robot by Ogata et al. [52], described by users as "tame and likable" that possesses an emotion-like model of an internal secretion system with states like radical unpleasantness, unpleasantness and pleasantness;
- A museum tour-guide robot Minerva, where ad hoc emotional states, ranging from happy to angry, served to improve navigation in crowded environments [53, 54];
- A socially interactive robot head ERWIN, in which five basic emotions are generated through modulation of hormonal-like parameters [55];
- Robovie-mini R2 and Robovie M by Nakagawa et al. [56], which are provided with a method to control affective nuances by mapping dimensions of valence and arousal onto velocity and extensiveness of motion and body posture;
- Hanson Robotics android head "Einstein" [57], which is capable of learning and producing a large number of realistic facial expressions based on Ekman's Facial Action Coding System, FACS [41];
- Huggable robot Probo by Goris et al. [58] capable of producing emotional expressions based on the circumplex model of affect on a;
- Expressive robotic head EDDIE [59] capable of display of affect based on the circumplex model and Ekman's FACS;
- Robotic dog AIBO programmed to express Extraversion and Introversion [60];
- CERO, a fetch-and-carry robot for motion impaired users [61].

Finally, in addition to AIBO, a number of other commercial robots designed for Human-Robot Interaction, such as Sony's QRIO [62], IBM's QB [63] and NEC's PaPeRo [64], claim to possess some level of affective capabilities.

The systems presented in this subsection have a very limited affective component to them, primarily confined to emotions. They also are mostly engineering-oriented, without psychological or biological foundations. They are described here for completeness and to provide an idea of the extensiveness of the field.

2.2 EVALUATING SOCIALLY INTERACTIVE ROBOTS

Despite the fact that the field of human-robot interaction in general, and socially interactive robots in particular, is quite young, a number of researchers have conducted studies where human attitudes towards robots expressing personality or emotion were assessed. The largest study to date was conducted by Bethel et al. [65], in which 128 participants interacted with two different non-anthropomorphic search-and-rescue robots in a mixed design. This study utilized four methods of evaluation (self-assessments, video-recorded observations, psychophysiology measurements, and a structured audio-recorded interview) so that convergent validity could be obtained to determine the effectiveness of the use of non-facial and non-verbal affective expression for naturalistic social interaction in a simulated disaster application. The results indicated that participants were calmer in the emotive mode, and reported feeling that the emotive robots were friendlier and spent more time oriented towards them [66]. Based on this and a review of other HRI studies, Bethel et al. [65] also provide a number of recommendations for design, execution and analysis of these types of experiments, stressing, in particular, the importance of large sample sizes and multiple evaluation methods.

The rest of this subsection describes a number of representative studies conducted in the area. In particular, Yan et al [60] encoded expressions of Introversion and Extraversion in AIBO and found that subjects could correctly identify the encoded trait, preferring to interact with a robot possessing a complementary personality. In a set of studies by a different group [67], it was observed that the personality preference depended on the nature of the task given to the subjects, and that people liked more cheerful robots better, but followed a serious robot's instructions to a greater extent. In a similar study performed by Mutlu et al. [68] it was found that the task structure (cooperative vs. competitive) and user attributes (male vs. female) impacted users' perception of the robot; an additional finding from the same study suggested that metrics used in human-robot interaction studies do not always have the same validity as in human-human study, and should be carefully reexamined or new metrics should be developed.

In Lisetti et al. [36], the user study followed a social informatics approach, according to which experiments are conducted in order to inform the design of the robot. It was found that after being exposed to Cherry, the aforementioned office robot possessing emotions and personality, the participants responded more positively towards robots with social abilities, and in particular towards those expressing emotion, both positive and negative. In the study by Bruce et al. [69], interaction with Vikia, a robot with an on-screen woman's face capable of emotional expression, was compared to one without a face, and it was found that more people were willing to stop and interact with the robot with the face than without one.

Other interesting findings come from studies with Sony AIBO. For example, Kahn et al. [70] conducted a study in which 80 children were engaged in 45 minute individual sessions, where half of each session they interacted with AIBO, and the other half with a stuffed dog. No difference in subjective evaluation between the two dogs (including "having feelings") were found, but there were behavioral differences: children engaged more in exploratory and apprehensive behaviors, as well as attempts at reciprocity with AIBO, and more often mistreated the stuffed dog. In another study, Friedman et al. [71] analyzed over 3,000 postings on on-line discussion forums from 182 participants that had something to say about AIBO directly, and found that 38% of members spoke of AIBO's having feelings, 28% spoke of emotional connection (mostly person to AIBO), and 26% spoke of companionship with AIBO. However, AIBO's moral standing (e.g, deserving respect and having rights) was very low (12%).

In addition to evaluating social robots for more or less general purposes, a number of studies have been conducted in order to assess effects of such robots on human well-being. In fact, a new term has been coined – “robopsychology”, defined by Libin et al. [72] as a systematic study of compatibility between people and artificial creatures on many different levels, such as sensorimotor, emotional, cognitive, and social. As an evaluation tool for robopsychology, they developed a Person-Robot Interactive Scale - A Unified Methodology for Assessing Interactions Between Humans and Artificial Creatures. It includes nonverbal interaction scale (tactile and manipulative), verbal interaction scale, emotional display scale, and the animated interaction scale. Using this scale for assessment, Libin et al. [72] conducted a study of 80 people interactive with a robotic cat NeCoRo [73] to show how individual and social differences influence the interaction between a person and a robot; one interesting finding was that older people liked the interaction better. Unfortunately, no further developments in this research direction could be found.

Another study that falls under this category is one assessing human interaction with the aforementioned seal robot Paro at a day service center [44]. According to the study, the interaction improved the mood of elderly subjects, their perceived vigor, and stress recovery; also, according to the observations of the nursing staff, the subjects were more active and communicative during the interactions. Finally, a number of studies have been done to analyze the interaction between robots and autistic children, both short term and longitudinal (for one day a week, 30-40 minutes per session, over a five week period) [74, 75].

To conclude this subsection, a number of assessment techniques used by researchers in evaluating interaction with social robots are described briefly:

- Dautenhahn et al. [74, 76] propose a combination of quantitative and qualitative techniques. The quantitative approach is based on analysis of micro-behaviors presented as well-identifiable, rather low-level and action/movement oriented categories, e.g., touch, eye contact, handling, approach, etc. The qualitative approach is based on conversation analysis (CA) and provides a systematic analysis of everyday and institutional talk-in interaction [77].

- Kanda et al. [78], in assessment of their socially interactive "partner" robot Robovie, used a combination of subjective evaluation by means of semantic differential with measurement of body movement; it was found that the body movement (e.g., distance, distance moved, eye contact, touch, etc.) was correlated with subjective evaluations.

- Scholtz et al. [79] developed a set of metrics for bystander social interaction with robots: predictability of behavior, capability awareness, interaction awareness and user satisfaction.

- Bartneck et al. [80] present a survey of subjective evaluation measures used in a number of HRI studies, and provides a number of semantic differential scales to evaluate the following constructs: anthropomorphism, animacy, likeability, perceived intelligence and perceived safety of robots.

- Along with design recommendations, Bethel et al. [65] also provide an general overview of evaluation methods advisable for use in HRI studies, including self-assessments, behavioral measures, psychophysiology measures, interviews and task performance metrics.

Although most of the studies described here do not evaluate affective robotic behavior directly, but rather socially interactive robots in general, the evaluation methods utilized appear to be applicable to evaluating human interactions with affective robots as well. Such traditional methods as self-assessment and observation can be applied to understand participants' perceptions of affective robots, and should be combined with more objective techniques, such as task performance or compliance metrics, to provide greater validity.

2.3 AFFECT FOR ENHANCING ROBOTIC BEHAVIOR

For a limited number of robotic systems, improving human-robot interaction was not the primary reason for incorporating affect. In these, the role of affect as an adaptive mechanism was acknowledged and therefore the focus is on improving the robot's behavior through decision making, learning, action selection, and multi-robot cooperation. However, it should be noted that none of these systems employ a combination of affective phenomena; rather, they are limited to emotions as their only affective components.

Velasquez extends the role of emotion from emotional expression for communication purposes to a determining factor in decision-making processes [81]. His model includes multiple mechanisms of emotion generation, based on Izard's [82]'s four types of elicitors of emotion in humans: neural, sensorimotor, motivational and cognitive. This approach is categorical, where the affect space is divided into a number of distinct emotions. The model can synthesize a number of emotions simultaneously, and allows for a number of different affective behaviors to be active at once. Decision-making is emotionally biased not only by the current emotional state, but also by prior emotional experiences, based on Damasio's [9] somatic marker hypothesis. Velasquez also considered temperament and mood as potential influences on emotion generation. This model was implemented in Yuppy, an emotional pet robot, and Virtual Yuppy, its simulated counterpart.

A number of researchers used the Fuzzy Logic approach to emotion generation. In El-Nasr et al.'s emotional system for decision making in mobile robots [83], the emotional states have no definite boundaries, and are represented by fuzzy sets with intensities of low, medium and high. They are generated according to Fuzzy Logic inference rules and OCC model [19], based on goals and expectations, where at different intensities the same emotion can trigger different actions. In another fuzzy-logic based system [84], Fuzzy Cognitive Maps (FSM) are used to represent generation and effects of emotional states. FSMs allow a robot to learn associations between stimuli and emotional states, as well as between emotions and tasks. Finally, Yu et al. [85] 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 sensor input and current emotional history and can influence behavior selection by increasing/decreasing corresponding action weights.

Murphy et a. [86] 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, as in the aforementioned work on robot Cherry [36]. The Emotional State Generator (Finite State Machine) accepts measures of task progress as input, and emotion then influences 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.

In the Schema-Based Agent Architecture by Scheutz [87], basic motivations and emotions are also used as control mechanisms. The architecture is based on Arbib's schema theory [88], and consists of sensors, perceptual schemas, motor schemas, with an action selection mechanism through summation and effectors. The emotion schemas are placed in the intermediate position between perceptual and motor schemas; they take perceptual stimuli as input, and change "behavioral dispositions" of agents by influencing behavioral gains, thus allowing agents to adapt their behaviors. A basic mammalian fear/anger system was implemented in the architecture in which controllers were modeled as a feed-forward, three-layer neural network. It was found in a set of simulation experiments that agents with emotional control mechanisms performed much better in a variety of foraging and survival tasks.

Another control architecture for autonomous mobile robots which incorporates biologically-inspired artificial emotions is proposed by Lee-Johnson et al. [89]. 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.

Dominguez et al. [90] present a Real-Time Emotional Architecture (RTEA), which regulates robotic behavior to fulfill objectives depending on emotional state, taking into a consideration robot's attitude (determines appraisal styles for emotion generation) and robot's mental capacity (processing resources). Positive and negative emotional responses are generated based on the environmental conditions and robot's state, and their influence on behavior can be inhibited if necessary. This model appears to be primarily engineering-driven, and the attitudes in particular differ significantly from more widely accepted definitions, and resemble personality more than anything else.

Other systems aimed at enhancing overall robot performance include those employing hormonal homeostatic regulation to guide consummatory and appetitive behaviors [91]; those using a global emotional state to help regulate recharging needs of autonomous robots [92]; those guiding reinforcement learning through maintaining homeostatic variables [93]; and those establishing general goals through emotions based on both current and future well-being of the robot [94].

Although none of the systems in this subsection are directly related to the research in this proposal, as they only include one affective phenomenon – emotions (with the exception of RHEA, which includes "attitudes" as well), they are described here because they all claim that emotions may enhance general robotic behavior, and therefore are of interest to our overall research goal.

2.4 AFFECT IN VIRTUAL AGENTS

The class of "virtual agents" includes a wide range of agents capable of facial and/or bodily behaviors and created for a variety of purposes, such as education/training, entertainment, and commerce. The underlying premise behind these agents, including embodied conversational actors, autonomous virtual characters and synthetic agents for interactive drama, is believability. A believable character can be defined as "one who seems lifelike, whose actions make sense, who allows you to suspend disbelief" [95].Therefore user perceptions are the key, and play a paramount role in design of these characters. In particular, intentionality is believed to be very important in humans, and so an overwhelming majority of the systems in this domain are highly goal-oriented and cognitively inspired. It is no wonder then that the appraisal-based OCC model of emotion by Ortony et al. [19] is at the core of the affect modules. The OCC model classifies emotion according to cognitive eliciting conditions, rather than using sets of basic emotions, and includes a rule-based system for generation of emotion types. As was the case in the previous section, the most relevant work is presented first, and then a brief overview of other affect-incorporating systems is given.

2.4.1 COMBINATION OF PERSONALITY, EMOTIONS AND MOODS

The most relevant to our research system is perhaps the one proposed by El Jed Mehdi et al. [96]. Not only does it model three of the four components proposed for *TAME*, it also provides a number of interactions between these components, and the psychological foundations behind the personality traits included in this system are similar to the ones proposed for this research. This model is based on the generic model for personality and emotion developed in MIRALab [97]. The two systems have different purposes: the former is aimed at allowing the user to animate virtual characters and provides the underlying emotional background for this process, whereas the latter is used as glue for perception, dialogue and expression. Both models consist of emotions represented as an m-dimensional vector of intensities that dynamically change, personality traits represented as a static n-dimensional vector intensities, and mood represented as a dynamically changing k-dimensional vector of intensities. The personality traits used are based on the Five- Factor Model and influence emotion generation by changing sensitivity of a character to certain emotions. The emotion emergence is based on the OCC model, and depends on the current situation, previous emotional state, and input from characters personality and current mood; El Jed Mehdi et al. [96] also add a decay function into the mix. The mood state is based on the current emotion experienced and the personality of the character. The output of the Emotion module in El Jed Mehdi et al. [96] affects the facial expression of the character, its behavioral reaction (emotional gesture) and cognitive reaction. For example, the same character exhibits a different walking style as a result of an emotional response. In Egges et al. [97], emotional response results in affective facial expressions and changes in dialogue structure. Unlike the framework in this proposal, these models do not consider personality and moods apart from their influence on emotions, do not explore time-varying aspects of these phenomena, and have not been considered for the robotics domain.

2.4.2 COMBINATION OF PERSONALITY, EMOTIONS AND ATTITUDES

Prendinger et al. [98] present the SCREAM system – a scripting tool that enables authors to create emotionally and socially appropriate responses for animated characters. In this system a complex process for emotion generation (based on the OCC model), resolution, maintenance and regulation is employed which is heavily influenced by a character's personality and attitudes. For example, personality plays a role in selecting the dominant emotion among a number of conflicting ones, determines the decay process for emotions, and regulates emotional expression (whether or not the dominant emotion will be acted upon). One of the most relevant parts of this system, however, is the process of attitude change which is based on combination of emotions at the time of attitude formation and whether a strong attitude has been formed already and is either to be strengthened or weakened in the consequent interactions.

2.4.3 OTHER AFFECTIVE SYSTEMS

The OCC cognitive model also serves as psychological foundation for a number of non-human animated characters. One example from the entertainment domain are believable agents from the "Oz project" by Joe Bates, Bryan Loyall and Scott Reilly [99]. These animated creatures are capable of producing a range of emotions and corresponding behaviors. The emotion generation system "Em" is based on the OCC model, and emphasizes cognitive appraisal for emotion generation. In this system, intensity of emotions depends on the importance of the goals that generated them. An emotion has to reach a certain threshold before it can activate a corresponding behavior. If negative emotions dominate, then the agent is in a 'bad mood', and vice versa. Em was embodied into an architecture for Action, Emotion, and Social Behavior, in which the output of Em influenced the behavior by altering the importance of current goals. The architecture also includes "behavioral features", such as aggressive and curious, which affect the mode of execution of actions.

Another OCC-based system is Clark Elliot's "Affective Reasoner" [100] which simulates simple worlds populated with agents capable of responding emotionally as a function of their concerns. In this system, Elliot models three categories of emotion intensity variables: situation-event variables (independent of situation interpretation), stable disposition variables (involved in an agent's interpretation of the situations, which are constant and help to determine an agent's personality), and mood-related variables. All three categories affect the intensity of emotions rather than influence behavior directly.

Other systems incorporating affect into believable characters include:

- A domain-independent framework for emotion and adaptation (EMA) by Gratch et al. [101] and Marsella et al. [102] which includes appraisal-based emotions, as well as mood expressed as an aggregate emotional state;
- A Dynamic Belief Network-based Greta's Mind [103], in which OCC-based emotions are influenced by agent's personality;
- P-R Planning Architecture in which personality is modeled as a cluster of goals [104];

- MAMID architecture which models both static (traits) and dynamic (affective state) processes via a combination of belief networks and rule based inferencing; a dimension-based Bayesian model of personality and emotion by [105];
- An interactive AlphaWolf system in which dimension-based emotions and emotional memories help maintain the believability of animated wolf pups while giving the high-level behavioral control to the user [106].

2.5 SUMMARY

Table 2 provides an overview of the aforementioned affective systems modeling more than one type of affect, including each system's name, the affective components it contains, and the primary domain (human-robot interaction, autonomous robots, or virtual agents) it is designed for.

Table 2: Overview of Multi-Affective Systems

System	Affect Types	Domain
Roboceptionist	Emotions, Moods, Attitudes	HRI
Waseda Eye No 4 Refined	Emotions, Moods, Personality (with respect to Emotions)	HRI
Cherry	Emotions, Personality (with respect to Emotions), Moods	HRI
Character Robot Face	Emotions, Moods	HRI
iCat	Emotions, Moods	HRI
Paro	Emotions, Moods	HRI
Velasquez's Yuppy Robot	Emotions; potentially temperament and mood	Autonomous Robots
Real-Time Emotional Architecture	Emotions, Attitudes	Autonomous Robots
El Jed Mehdi's Agent	Emotions, Moods, Personality	Virtual Agents
MIRALab's Agent	Emotions, Moods, Personality	Virtual Agents

Table 2 (continued)

SCREAM system	Emotions, Attitudes, Personality (with respect to Emotions)	Virtual Agents
"Oz Project"	Emotions, Personality	Virtual Agents
"Affective Reasoner"	Emotions; Moods and Personality as Emotion-generation variables	Virtual Agents
EMA system	Emotions, Moods	Virtual Agents
Greta's Mind	Emotions, Personality	Virtual Agents
MAMID Architecture	Emotions, Personality	Virtual Agents

Although all the presented systems incorporate affect on some level, they all differ from the proposed research in a number of ways. First, none of the systems includes all four of the proposed components, namely, personality, attitudes, moods and emotions. Furthermore, these systems are not exploring and comparing the contributions of each of these phenomena individually and in their time-varying interaction with one another. Finally, the use of affect in these systems seems to either be targeted towards improving social interaction with humans, or towards enhancing general robotic behavior, but not both, while this research covers both domains.

3 ARCHITECTURAL FRAMEWORK

Despite an increased interest among psychologists in the “feeling” part of our lives, there as yet exists no single unified theory of affect-related phenomena. Partly, the current state of things is due to the fact that the study of affect is still in its nascence and the existing body of knowledge regarding each type of affect has not reached the critical mass that would allow for comprehensive, all-encompassing and coherent theories that would satisfy the majority of researchers. Furthermore, traditionally, Emotion Psychology, Personality Psychology and Attitudes Psychology have been treated as separate fields, and although their mutual influence on each other has been acknowledged, little has been done to date to unify this type of research. Nonetheless, the existing work in these areas has a lot to contribute and can serve as an inspiration for endowing autonomous robots with adaptation capabilities, as well as improving their performance in the area of human-robot interaction. Each of the four components performs a distinct adaptive role: personality traits serve as an adaptation mechanism to specialized tasks and environments, emotions mobilize the organism to provide a fast response to significant environmental stimuli, moods bias behavior according to favorable/unfavorable environmental conditions, and attitudes guide behavior towards desirable goals and away from aversive objects. This said, however, it should be noted that although this framework is inspired by a number of separate psychological theories, it is not intended to be a cognitive model of affect and personality, but rather to serve as a framework for including affect into behavior-based autonomous robotic systems. In addition to robotics, the framework can also be extended to the field of intelligent embodied agents without loss of generality.

In this chapter, the first research subquestion, *“How can these affective phenomena be modeled in a robotic system?”*, is explored in detail by presenting the following:

- The psychological and cognitive foundations behind each component modeled in the *TAME* framework (Traits, Attitudes, Moods, Emotions);
- Computational representation of each component;
- Antecedents and methods of generation for each affective phenomenon;
- Their influence on robotic behavior;
- And interrelation between the components.

The primary focus of this chapter is on the psychological foundations and a computational model of the *TAME* framework; the software design incorporating the computational model and its implementation are addressed in Chapter 5.

3.1 TAME FRAMEWORK OVERVIEW

The goal behind the *TAME* framework [107, 108] is to add an affective element to a behavior-based robotic system in order to improve general robotic performance and facilitate human-robot interaction. In the behavior-based paradigm, a robot's control program consists of a collection of behaviors and coordination mechanisms [109]. Primitive behaviors have a set of defining parameters (e.g., obstacle avoidance sphere-of-influence) and these behaviors can themselves be combined into behavioral assemblages, where each of the primitive behaviors' outputs is weighted and combined, resulting in coherent motor actions. Perceptual input not only serves as stimuli for behaviors, but also triggers transitions between assemblages.

The affective module of the *TAME* framework is composed of four interrelated components: personality Traits, affective Attitudes, Moods, and Emotions. The input into this architectural module consists of relevant perceptual information, 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). Each component is modeled as a set of primitive behaviors and, with the exception of traits that are defined prior to execution, runs as a separate thread continuously throughout the execution. Instead of directly defining behavioral transitions, the affective module rather modifies the underlying behavioral parameters, which, in turn, directly affect currently active behaviors. As behavioral parameters can refer not only to primitive behaviors, but to complex assemblages as well, the module can, in effect, determine action selection, by setting, for example, the relative weight of one of the behaviors composing an assemblage to zero. The conceptual view of the framework is presented in **Figure 1**.

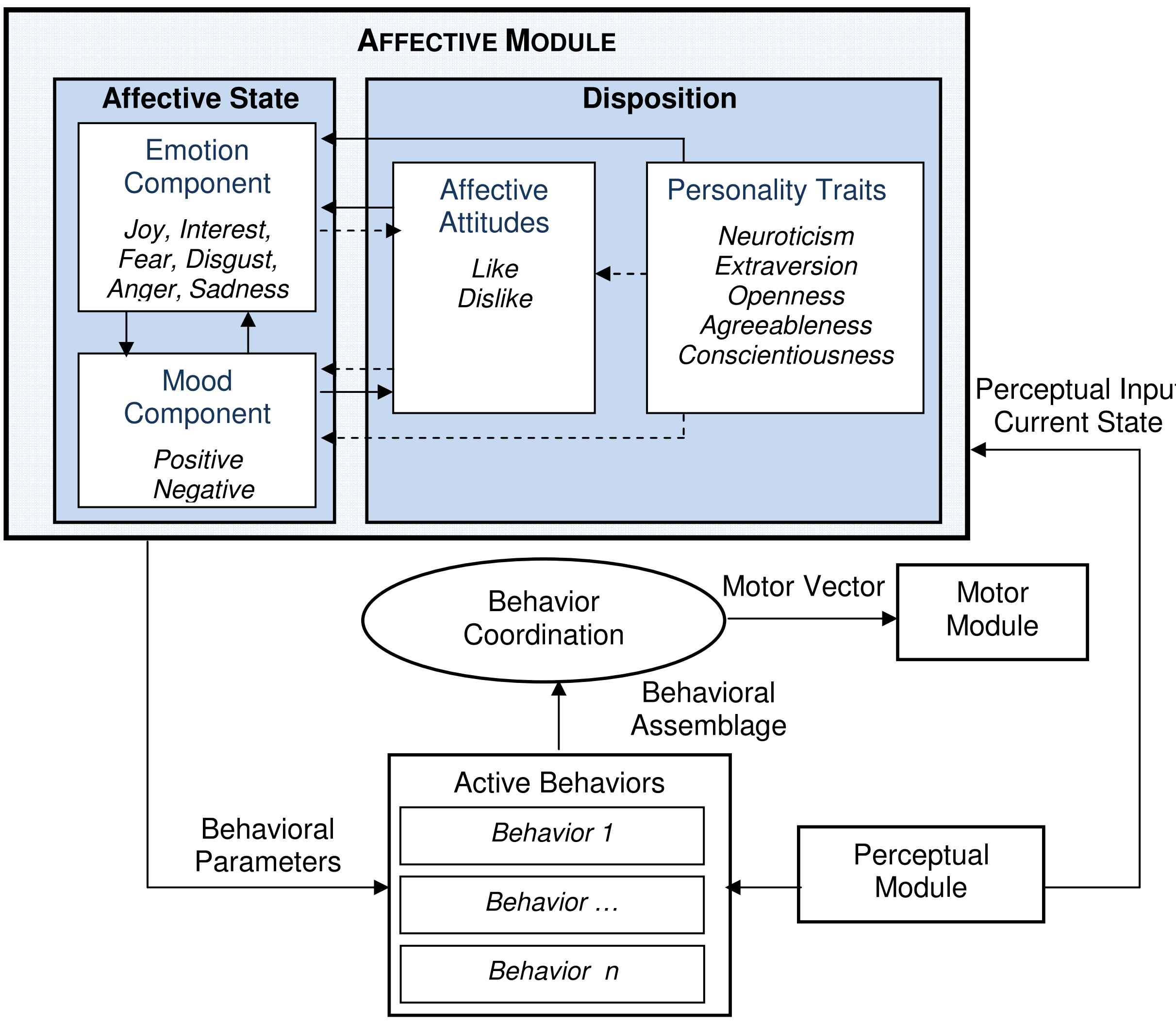


Figure 1: Integrative Framework for Affective Robotic Behavior (*TAME*) [107, 108, 110]. Dashed arrows represent potential interactions not currently explored in this research.

Affective state refers to temporary conditions of an organism reflected in multiple systems, whereas *disposition* reflects general tendencies to behave and process information in certain ways. 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" [10]. 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, or sentiments, are object-specific; however, unlike emotions, they refer to ways of seeing and treating an object rather than to momentary responses [3]. Finally, personality refers to enduring individual differences in behavior and information processing of a more general, object-independent kind.

Therefore, we can position each component in the two-dimensional space defined by duration and specificity [4, 7, 111]. Traits and emotions are at the opposite ends of the spectrum: traits are life-long, stable over time, and global (independent of specific objects/events), whereas emotions are short-term, dynamically changing and focused; moods and attitudes occupy the intermediate positions (**Figure 2**). In addition to occupying a different position in the duration/specificity space, each of the four components also performs a distinct adaptive role (not limited to what is described below): traits serve as an adaptation mechanism to specialized tasks and environments, emotions mobilize the organism to provide a fast response to significant environmental stimuli, moods bias behavior according to favorable/unfavorable environmental conditions, and attitudes guide behavior towards desirable goals and away from aversive objects, as well as facilitate decision-making process by reducing decision space.

We could visualize the complex variation in time and intensity of these components as a multi-dimensional surface, where traits would present a never-changing base plane, with moods representing low intensity, smooth undulations on this plane, and emotions being relatively rare, sharp, high-intensity and short-term spikes superimposed onto moods. Attitudes are harder to pinpoint, as they are both objects-specific and lasting; perhaps, they could be visualized as multiple lines running across the aforementioned surface.

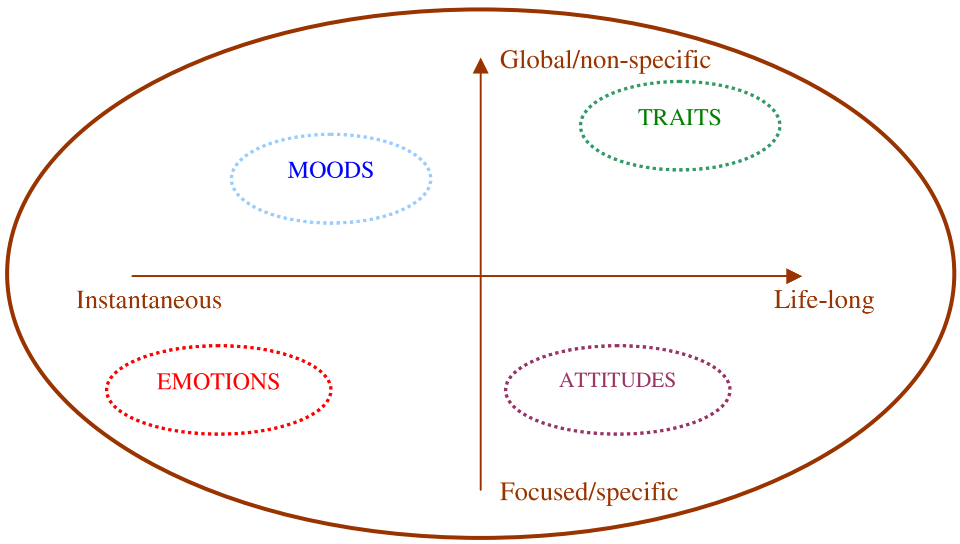


Figure 2: Relative Positions of Types of Affect in the Duration/Specificity Space

In the remainder of this chapter, we will discuss each of the components of the affective modules: their psychological foundations, representation, generation, and influence on behavior, and well as their interrelation with each other.

3.2 PERSONALITY TRAITS

In humans, personality is evident in a great variety of situations and affects numerous behaviors, as well as our emotional responses, our daily moods, and the way we form and express our attitudes. Personality traits identify the consistent, coherent patterns of behavior and affect that characterize individuals; they produce a profound influence on generation and application of affective phenomena. In *TAME*, this component also provides the widest influence on the robotic behavior of all other components.

3.2.1 PSYCHOLOGICAL FOUNDATIONS

Personality psychology covers a wide range of psychological phenomena, and attempts to combine this multitude of phenomena into a unifying theory, both on the level of human nature universals, and on the level of individual differences [112]. Among many directions of personality research, the study of personality traits seeks to identify the consistent, coherent patterns in behavior, or behavioral dispositions, that characterize different individuals and to classify them according to a number of broad dimensions. From the evolutionary standpoint, individual differences might have resulted from adapting to certain types of environments: individuals, by being different, can select environmental niches that maximize their own fitness [6]. Matthews et al. [113] also suggest that "traits represent adaptation to specialized environments, defined by their processing requirements, and are supported by a number of independent cognitive characteristics" (e.g., neurotic individuals are fit for environments with subtle threats), and thus have a purpose and a cognitive foundation. In the next section we will review the "Big Five" taxonomy of personality traits, the most prominent and comprehensive of trait models as of today.

3.2.1.1 The Big Five Trait Taxonomy

In the late 1980s – early 1990s, two research groups independently arrived at a set of five global personality dimensions via factor analysis. The taxonomy developed by McCrae et al. [114] was named "The Five-Factor Model of Personality" (FFM), and the one developed by Goldberg [115] - the Big Five. Although not without their differences, both taxonomies are similar enough to be treated interchangeably for the remainder of the proposal, and will be referred to as "the Big Five" or "Five-Factor Model".

The Big Five taxonomy represents diverse systems of personality description in a common framework [116]. Its five dimensions (domains) are: *Openness, Agreeableness, Conscientiousness, Extraversion, and Neuroticism.* The taxonomy is consistent over time, age and cultural differences [114], as well as applicable to nonhuman animals [117]. Personality traits, according to McCrae et al. [114], are mainly inherited or imprinted by early experience, and remain relatively unchanged throughout the lifetime. Numerous studies also show predictive utilities of these dimensions; for example, Conscientiousness, a domain of Big Five, is a general predictor of job performance, Extraversion predicts performance in sales and management positions, Neuroticism can show vulnerability to depression, and Openness is a predictor of creative performance [116].

Given the broad nature of these dimensions, their labels can be somewhat misleading, therefore in **Table 3** we present the most commonly used names along with a brief description of each dimension. In addition, we will also include a number of facets per each dimension identified as more specific constituents of global dimensions [118].

Table 3: Description and Facets of Big Five Personality Dimensions

Big Five Dimension	**Description**	**Facets [118]**
Extraversion/ Energy	Implies an energetic approach to the social and material world [116]; also refers to liking people and preferring large groups and gatherings [114].	• Gregariousness (sociable) • Assertiveness (forceful) • Activity (energetic) • Excitement-seeking (adventurous) • Positive Emotions (enthusiastic) • Warmth (outgoing)

Table 3 (continued)

Agreeableness/ Affection	Contrasts prosocial and communal orientation toward others with antagonism [116]; a dimension of interpersonal tendencies, refers to being altruistic, sympathetic to others, cooperative and eager to help [114].	• Trust (forgiving) • Straightforwardness (not demanding) • Altruism (warm) • Compliance (not stubborn) • Modesty (not show-off) • Tender-mindedness (sympathetic)
Conscientiousnes/ Control	Describes socially prescribed impulse control that facilitates task- and goal-directed behavior, such as following norms and rules, and planning, organizing and prioritizing tasks [116]; high scores mean purposeful, strong-willed, achievement – oriented individuals [114].	• Competence (efficient) • Order (organized) • Dutifulness (not careless) • Achievement striving (thorough) • Self-discipline (not lazy) • Deliberation (not impulsive)
Neuroticism/ Negative Affectivity	Contrasts emotional stability and even-temperedness with negative emotionality [116]; the general tendency to experience negative affects such as fear, sadness, nervousness, anxiety, etc. [114].	• Anxiety (tense) • Angry hostility (irritable) • Depression (not contented) • Self-consciousness (shy) • Impulsiveness (moody) • Vulnerability (not self-confident)
Openness to Experience/ Open-mindedness	Describes the breadth, depth, originality and complexity of individuals mental and experiential life [116]; refers to active imagination, preference for variety, intellectual curiosity, and independence of judgment [114].	• Ideas (curious) • Fantasy (imaginative) • Aesthetics (artistic) • Actions (wide interests) • Feelings (excitable) • Values (unconventional)

3.2.1.2 Implications for TAME

Based on this psychological overview, there are a number of design decisions we can make:

1. As traits are to a great extent determined by heredity and early experience, we will assume that their influence doesn't change throughout a "life-time" of a robot (be it a single mission, a general task over a number of executions, or all the interactions with the same person). Therefore, they can be set once at the beginning of the execution by an operator/designer, and will continually influence the selected behavioral and affective parameters.

2. As the trait component doesn't change throughout execution, it will accept no input, including the influence of other affective components; its output will consist of behavioral parameters that will serve as the base/default parameters.

3. As traits are in part a mechanism for adapting to a variety of situations, different tasks may require a different trait configuration that will affect the robotic behavior in different ways. As part of this research, we will provide a mechanism by which traits will affect actions. Although only the five broad dimensions will be used initially, the component may be potentially extended to explore their facets as well.

4. In addition to its influence on behavior, the trait component will also affect generation, intensity or response output of other affective components (described in later sections).

3.2.1.3 Application to HRI

The goal of including personality traits into a robotic system is at least two-fold: to adapt behaviors to different tasks and environments, and to provide a means for matching a robot's personality to a human's personality for a better fit in prolonged, everyday interactions. Additionally, traits serve a predictive purpose, allowing humans to understand and infer the robot's behavior better; for example, an extraverted robot would be expected to be more gregarious, loud and interactive than its introverted counterpart.

Table 4 presents examples of tasks and situations for which each of the traits would be beneficial, along with the general role the robot would have in each of the examples.

Table 4: Examples of Traits Applied to Human-Robot Interaction

Trait/Dimension	Example	Role
Extraversion	Extraverted robot would be more suitable for tasks requiring engagement and entertainment – for example, a museum tour guide, or a play partner for children; on the other hand, introverted robot would be more suitable for tasks requiring concentration from a human – any mutual problem solving task.	*Co-worker* *Co-inhabitant*
Agreeableness	A highly agreeable, compassionate robot would be valuable for working with sick and the elderly, where compassion is desired, whereas a more selfish robot might fare better in environments where robot abuse is expected (in case of non-acceptance)	*Caretaker* *Co-inhabitant*
Conscientiousness	A highly conscientious robot would be needed for tasks that required utmost precision and control – explosive material disposal; and a less conscientious robot would be less annoying for tasks in which time to completion is valued more than perfectionism – cleaning the house, for instance.	*Co-protector* *Co-inhabitant*
Neuroticism	A more neurotic robot would be suitable for dangerous environments, where it could suggest to an accompanying human, through its behavior, to be more vigilant and to pay more attention to any signs of danger; on the other hand, a more bold and calm robot may instill more confidence in the surroundings.	*Co-protector*
Openness to Experience	A robot high on Openness would suit collaborative exploration tasks, where the robot would favor more creative exploration strategies – space exploration or playing with children; whereas a less open robot would be preferred in more mundane cases, where just getting the job done is more important – cleaning or delivery tasks.	Co-explorer Co-worker

Similarly, combinations of different traits can be put together to match a task or an environment: for instance, an extraverted, agreeable (compassionate) robot would perform well as a nursebot in a recovery ward; and an introverted, less agreeable (stern) robot would project a sense of authority when managing a disaster evacuation route.

3.2.2 REPRESENTATION AND INFLUENCE ON BEHAVIOR

Traits are represented as a vector $\vec{p}$ of intensities of N personality traits, where intensity refers to the extent to which each trait is represented in the robot:

$$\vec{p} = [p_i]$$

For this research, we will use the five broad personality dimensions that constitute the Big Five taxonomy, therefore:

$$\vec{p} = \begin{bmatrix} Openness(O) \\ Conscientiousness(C) \\ Extraversion(E) \\ Agreeableness(A) \\ Neuroticism(N) \end{bmatrix}$$

3.2.2.1 Grounding Trait Intensity in Psychological Data

Although in theory trait intensity could range from negative infinity to positive infinity, it would be highly unlikely or useful. What would it mean to have a trait of certain intensity, either for humans or for robots? One way to ground it in both psychological findings and in reality is to look at the normal distribution of personality scores. Multiple tests have been developed to assess the relative personalities of both people and animals. According to one of such tests [118], the Five-Factor Model self-reported personality scores of 500 men and 500 women were normally distributed, with the following means and standard deviations (**Table 5**):

Table 5: Means and Standard Deviations for Personality Scores for Adults

Personality Trait	Mean (μ)	Standard Deviation (σ)
Openness	110.6	17.3
Conscientiousness	123.1	17.6
Extraversion	109.4	18.4
Agreeableness	124.3	15.8
Neuroticism	79.1	21.2

Given these scores, we can identify a range within which a certain percentage of the population falls. For example, approximately 68 % of the population falls within +/- 1 standard deviation of the mean, 95% within +/- 1.96 standard deviations from the mean, and 99% within 2.58 standard deviations from the mean. **Figure 3** displays the normal distribution curve of personality scores for Neuroticism, with the corresponding percentage/standard deviation points marked on the curve with black diamonds.

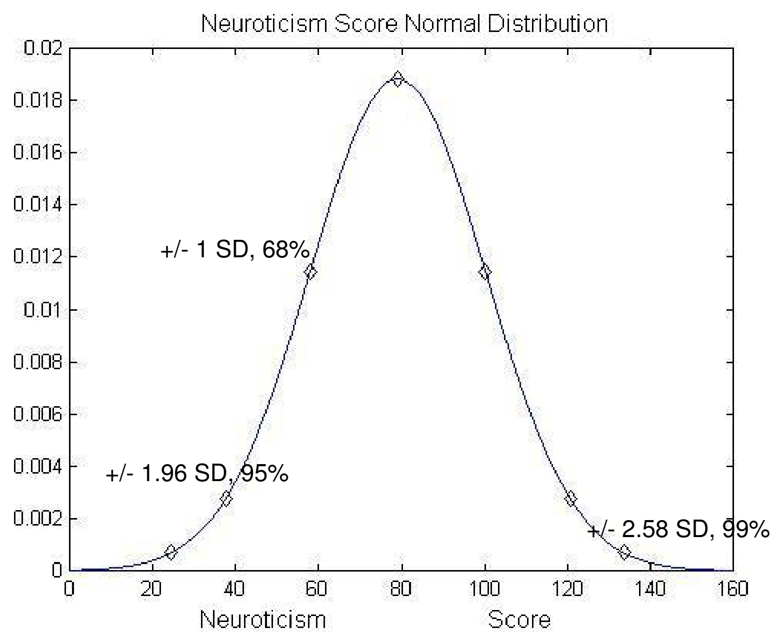


Figure 3: Neuroticism Scores Normal Distribution

The mean score for a trait refers to the average trait intensity; this might correspond to general suitability for a wide range of tasks, but not to any type in particular. As the values slide away from the mean, they become representative of a lesser proportion of a population, and also more adaptive for certain types of tasks, but not for the others. For example, a low value of Neuroticism signifies that an individual is calm, unemotional, and has low avoidance tendencies when compared to others; such an individual will thrive performing either high-risk tasks where courage is required or those with a low amount of danger.

In terms of robots, by using this bell curve as a baseline, we can assign to each trait numeric values that represent the relative location of the robot's desired personality against the average trait value. For example, to have a highly neurotic individual (person or robot) would mean that their neuroticism value would be greater than some large percentage of the population, e.g. 99%. We can then calculate mathematically the personality score that is greater than the personality scores of 99% of the population, or 134 (for Neuroticism). Therefore, prior to execution, the user defines a personality configuration, a value for each of the five traits, by selecting the desired population percentile that the robot's resulting personality should reflect. This population percentile is converted to a numeric value relative to the normal distribution for that trait by integrating over the probability density function of the normal distribution *N(μ,σ)*:

$$\begin{aligned} N(\mu_i, \sigma_i) &= \frac{1}{\sigma_i \sqrt{2\pi}} e^{-(x-\mu_i)^2 / 2\sigma_i^2} \\ K_i &= \int_{-\infty}^{p_i} N(\mu_i, \sigma_i)\, dx \end{aligned} \qquad (1)$$

where μ_i and σ_i are the mean and standard deviation of the personality trait being calculated, *N()* is the normal curve, K_i is the desired population percentile, and p_i is the corresponding personality trait value that is being solved for.

The advantage of using this method for assigning personality intensities, as opposed to simply selecting from a range of allowable values, is that it provides an operator with a more intuitive means for understanding what a certain numeric value stands for. In addition to providing us with the personality trait value corresponding to an arbitrary percentile, the same equation can also be used to estimate an upper and lower bound of interest. If 99% of the population should be within the desired range, then the lower bound for trait *i* will be at 0.5%, and will correspond to $\mu_i - 2.58 * \sigma_i$, and the upper bound will be at 99.5%, which corresponds to $\mu_i + 2.58 * \sigma_i$. For example, for 99% of the population centered about mean, 79.1, the lowest Neuroticism value will be:

$$lowest_{Neur} = \mu_{Neur} - 2.58 * \sigma_{Neur} = 79.1 - 2.58 * 21.2 = 134 \quad (2)$$

3.2.2.2 Determining Robotic Behavior using Personality Traits

As was mentioned earlier, different personalities may have resulted from adaptation to different kinds of environments and tasks; therefore it may be advantageous to provide robots with similar capabilities. For example, in a mobile robot, a higher level of Neuroticism may be expressed as exhibiting more prominent obstacle avoidance (i.e., keeping a greater distance between robot and obstacles). This would be suitable for more dangerous environments and in cases where the survival of the robot is more important than expeditious task performance. A number of behavioral parameters are involved in producing an obstacle avoidance response, including an obstacle avoidance gain, and obstacle avoidance sphere of influence. In general, we will define the set of all behavioral parameters that may be influenced by any one or more traits to be the behavioral space. In this section, as part of our research, we will investigate a functional mapping f(x) from the trait space to the behavioral space.

Initially, we use a polynomial mapping. In order to achieve such a mapping, two or more pairs of corresponding data points from both the trait and behavioral spaces are required. In particular, with 2 pairs of data points we can provide a linear correspondence by fitting them to a line, and 3 pairs of data points allow us to fit them to a 2nd degree polynomial. The default values of behavioral parameters provide a natural correspondence to the default (average) trait values. The second data point in behavioral parameters could be the lowest possible/desirable value of a particular parameter, with 0 being the simplest case (e.g., for the majority of parameters negative values don't make sense). For certain behaviors, the highest possible/desirable parameter values could also be obtained – e.g., sensor range would provide an upper limit for Obstacle Avoidance Sphere of Influence parameter. For gains, for example, the highest desired value would correspond to a value that would allow a particular behavior to dominate other currently active behaviors. By necessity, these lowest and highest possible/desirable parameter values will be obtained experimentally and might differ for different robotic platforms.

For the present, for parameters for which 3 data points are available, a second degree polynomial of the form $y = ax^2 + bx + c$ is advantageous, as it accounts better for cases where the default behavior lies closer either to the lowest or highest allowable parameter value (see **Figure 4** and section 3.2.2.2.1 for more detail.) For example, in the case where the default behavior lies closer to the lowest desirable parameter value, the mapping function will grow slower up to the default parameter value, and faster afterwards. The resulting trait-based behavior parameter will replace the default values: $B_{i,trait}=f_{ij}(p_j)$, where $B_{i,trait}$ is the new behavioral parameter, and $f_{ij}(p_j)$ is the functional mapping from the trait j to the behavioral parameter i.

In order to provide the mapping from each trait to the behavioral space, we need to know which parameters a trait affects, as well as the direction of its influence: direct or inverse. For example, the trait of Neuroticism is directly related to avoidance behavior, whereas a trait of Conscientiousness is related to such a behavioral tendency inversely (e.g., a more conscientious robot may risk bumping into obstacles more in favor of faster task completion). Therefore, a personality-behavior dependency matrix $\overrightarrow{pb} = [pb_{ij}]$ is defined, which specifies the presence/absence and direction of influence of personality traits on behavioral parameters: $pb_{ij} \in \{-1,0,1\}$, where 0 signifies absence of influence of trait *j* on behavior *i*, +1 means direct influence, and -1 is inverse influence. This matrix influences the creation of the functional mapping f_{ij}. In particular, the bounding trait values will be inversed for those traits that provide an inverse influence; and, if $pb_{ij} = 0$, then the mapping $f_{ij}(p_j) = 0$ for all values of p_j.

3.2.2.2.1 Obstacle Avoidance Example

Looking at an example of obstacle avoidance, the magnitude of the obstacle avoidance vector is defined as follows:

$$O_{magnitude} = \begin{cases} 0, & if \ d > S \\ \dfrac{S-d}{S-R}G, & if \ R < d \le S \\ \propto, & if \ \mathrm{d} \le \mathrm{R} \end{cases} \quad (3)$$

where *S* is the default sphere of influence, *R* is the radius of the obstacle, *G* is the default avoidance gain, and *d* is the distance of robot to center of obstacle.

The personality dimension of Neuroticism has been found to influence avoidance behavior [119], which in the case of a mobile robot may be expressed as a tendency to stay away from obstacles. The obstacle avoidance gain ***G*** is therefore directly affected by this trait: it grows as the value of Neuroticism increases. Once the new trait-based obstacle-avoidance gain is calculated through the functional mapping $f_{G,Neur}(p_{Neur})$, the default gain is replaced with the new G_{trait} value to calculate the magnitude of the obstacle avoidance vector for the duration of the robot's life-cycle (or task completion).

Let's suppose that there are two different configurations for default/lowest/highest parameter values, with both the lowest and the highest values kept constant ($lowest_{Obs} = 0.1, highest_{Obs} = 4$), but the default value is varying:

$default_{Obs,1} = 1, default_{Obs,2} = 2$ The result of fitting these data points to a 2nd degree polynomial is displayed in **Figure 4**. The Neuroticism values are plotted along the horizontal axis, and the resulting Obstacle Avoidance Gain values are plotted along the vertical axis. When the default parameter value is located as mid-point between the lowest and highest values, the mapping approaches a line. However, when the default value is closer to the lowest value, the result is more of a parabola, which provides for a smoother mapping between traits and behavioral parameters, avoiding discontinuities.

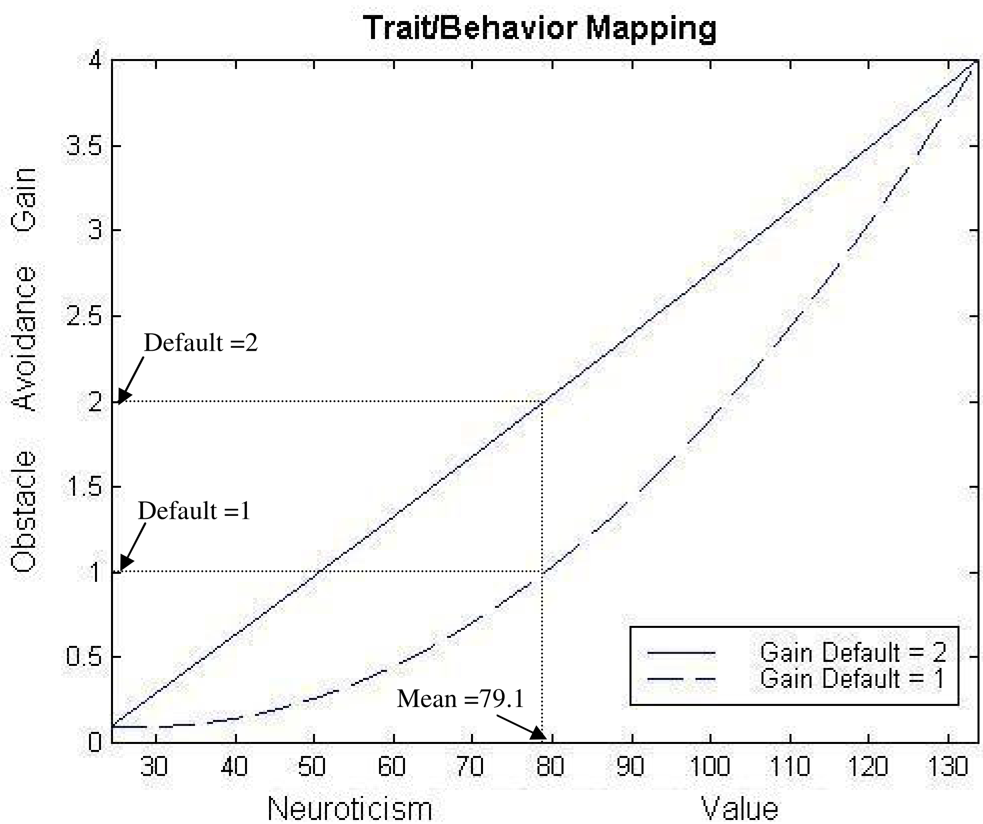


Figure 4: Mapping of Neuroticism to Obstacle Avoidance Gain

For the purposes of illustration, suppose that Neuroticism also inhibits the exploration tendency, that is, it produces an inverse influence on Wander gain. This gain is involved in producing Wander behavior, which generates a random direction vector, thus adding an exploration component to a robot's overall behavior. The Wander behavior has two parameters: random vector magnitude called "Noise_Gain"; and "Noise_Persistence" which controls the rate of directional switching of the vector. An example of both inverse and direct influences of Neuroticism on Wander gain and Obstacle Avoidance gain, correspondingly, is presented in **Figure 5** (the same desired upper and lower values for both behavioral parameters are shown for clarity). As we can see, as the value of Neuroticism grows the Wander gain becomes smaller (inverse influence), and vice versa for the Obstacle Avoidance gain.

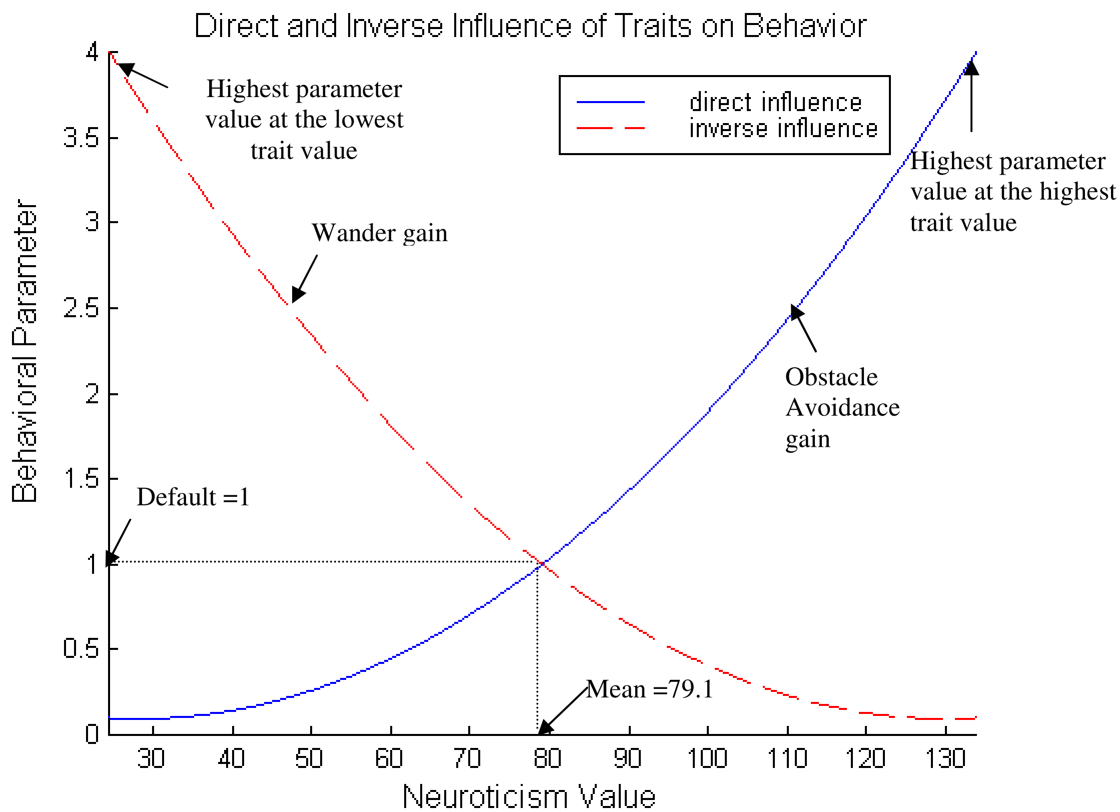


Figure 5: Comparison of Direct and Inverse Influences of Neuroticism on a Behavioral Parameter

3.2.2.3 Influence of Multiple Personality Traits on a Single Behavioral Parameter

Finally, each behavioral parameter may be affected by multiple traits. In such a case, first the trait/behavior mapping for each of the influencing traits is calculated according to the chosen function $f_{ij}(p_j)$, where trait *i* influences behavior *j*, a polynomial in this case.

Then, the results are averaged across all influencing personality traits to produce the final parameter value:

$$B_j = \frac{1}{\sum_{i=1}^{N} |pb_{ij}|} \sum_{i=1}^{N} f_{ij}(p_i) \qquad (4)$$

where B_j is a particular behavioral parameter, $f_{ij}(p_i)$ is the function that maps personality trait p_i to B_j, N is the total number of traits, and $\overrightarrow{pb}$ is personality/behavior dependency matrix. As was mentioned earlier, if there is no influence by trait j on behavior i, the result of $f_{ij} = 0$.

An example of multiple traits influencing the same parameter is given in **Figure 6**.

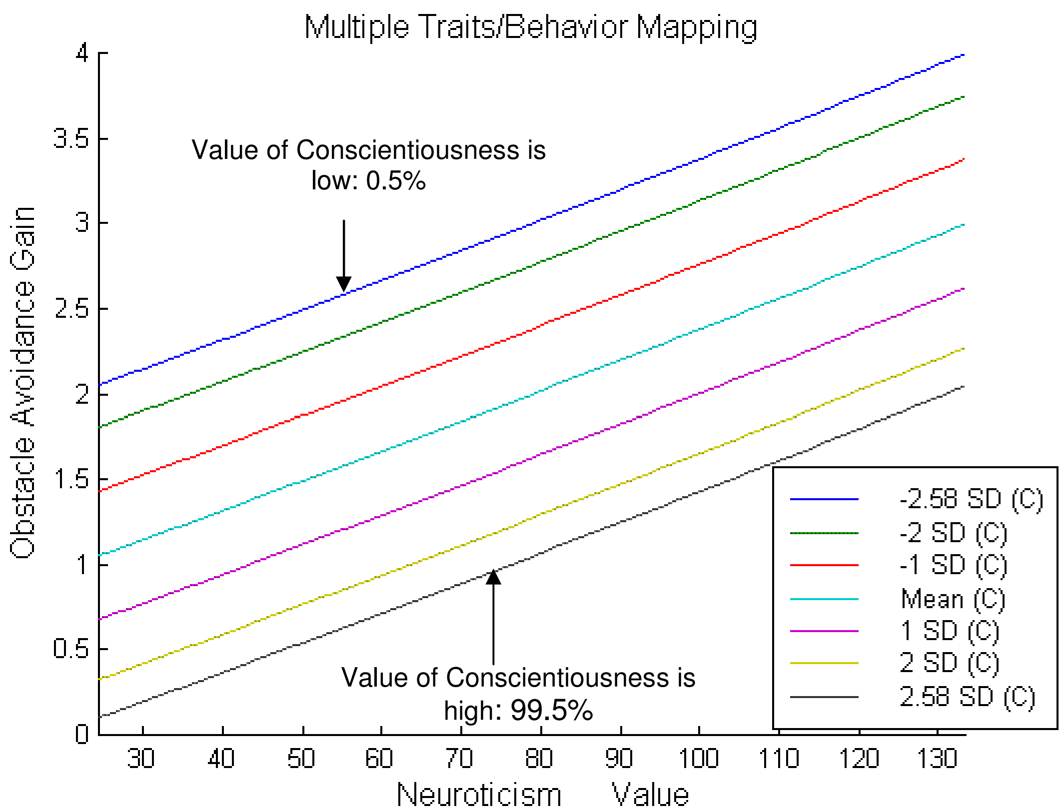


Figure 6: Obstacle Avoidance Gain as a Function of Both Neuroticism and Conscientiousness

The Neuroticism values are plotted along the horizontal axis, and the resulting Obstacle Avoidance Gain values are plotted along the vertical axis. The default, lowest and highest parameter values for Obstacle Avoidance gain are: $default_{Obs} = 2, lowest_{Obs} = 0.1, highest_{Obs} = 4$ The range of traits of Neuroticism and Conscientiousness is within 99%. The resulting Obstacle Avoidance gain is plotted for the following values of Conscientiousness: Mean, Mean +/- 1 SD, Mean +/- 1.96 SD (95 %) and Mean +/- 2.58 SD (99%). As shown in the figure, as the Conscientiousness score grows higher, the corresponding behavior parameter is lower, and vice versa for the Neuroticism values.

As trait values remain invariant throughout execution, the corresponding behavioral gains/parameters are computed only once at the beginning of execution. The resulting trait-defined parameters are now the base parameters upon which other affective components will provide their influence.

The same trait/behavior mechanism can be used for specifying expressive behavior for humanoid robots as well. In particular, given the minimum and maximum allowable angles for certain joints, and minimum and maximum desired frequency for some body movements and gestures, we can provide a mapping between personality traits and robot gestures, movements, and body posture. For example, an extraverted individual is characterized by more expansive and frequent gestures; therefore, high levels of extraversion would be mapped to more extreme joint angles (away from the body), and higher frequency, and vice versa for a introverted robot. Finally, in cases where it is not feasible to provide a smooth, safe trajectory for body movements, a number of variations of gestures/posture can be coded a priori (e.g., those representative of Low, Average, and High levels of Extraversion), and a simpler discretized mapping can be used (e.g., Average gestures for Mean +/- 1 SD, High for > 1 SD, and Low for < 1 SD).

3.2.3 TRAITS SUMMARY

In this section the psychological foundations for the trait component were presented, which serve as a basis for both translating user-defined trait combinations into an internal representation, and for defining a mapping from these configurations to behavioral parameters. In addition, potential applications of different traits to human-robot interaction have been discussed. Personality Traits is the only component in the *TAME* framework that does not vary over time, and therefore defines the baseline behavioral parameters.

3.3 EMOTIONS

"…our ignorance concerning emotions so grossly outweighs our knowledge."

Jaak Panksepp [120]

The construct of emotion has attracted significant interest in the area of computing in general, and robotics in particular, and is perhaps the most explored among other types of affect. Although our knowledge of emotions and their functionality is far from complete, as evidenced by the quote from Jaak Panksepp [120], we know enough to begin experimenting with a mechanism that has proven to be of a great adaptive value in mammals.

3.3.1 PSYCHOLOGICAL FOUNDATIONS

Given the complexity of the phenomenon, it should come as no surprise that there's a plethora of definitions of emotions [17, 121-124] – almost as many perhaps as there are emotion theorists. In different definitions, different aspects of emotions are emphasized – their adaptive significance, their components, the role of appraisal in emotion generation, etc. Some representative definitions can be found in **Table 6**. Although different, these definitions are not contradictory, and can be summarized with a set of points relevant for any autonomous agent, human or artificial. Emotions can be described as a coherent and organized mechanism or system that, under environmental contingencies, provides an organism with a fast, flexible, adaptive response by guiding, organizing, or coordinating behavior and perception.

Table 6: Representative definitions of Emotions

Researcher	Definition
Panksepp [124]	"… emotions are posited to be evoked under conditions having adaptational significance to the individual and to physically prepare and motivate the individual to contend with the adaptational implications of the eliciting situation."
Scherer [123]	Emotion is "an evolved, phylogenetically continuous mechanism that allows increasingly flexible adaptation to environmental contingencies by decoupling stimulus and response and thus creating a latency time for response optimization"
LeDoux [17]	"Basic building blocks of emotions are neural systems that mediate behavioral interactions with the environment, particularly behaviors that take care of fundamental problems of survival."
Johnson-Laird [122]	"… each of the discrete emotions serves distinct functions in the way it organizes perception, cognition, and actions (behavior) for coping and creative endeavors."

There are a number of features of emotions that should be considered while designing an artificial emotional system. Is there a set of distinct, separate emotions, each with its own distinct adaptive function, or can they be better described along a number of dimensions? What are the functions that emotions perform? Are emotions fast, hard-wired responses, or slower, more flexible cognitive ones? In the subsections below, we will touch upon these questions and try to identify the features that can help us in the consequent design.

3.3.1.1 Dimensional vs. Categorical Approaches

In the current psychological research, there are two main groups: those who define emotions not as separate mechanisms but as differing along a number of global dimensions (such as pleasant vs. unpleasant, low vs. high arousal, and approach vs.

avoidance), and those who contend that there exist distinct categories of emotions, each one with its own adaptive function which should be achieved from fundamentally distinct responses [125]. The theorists adhering to the dimensional view usually consider emotions as being socially learned and culturally variable [125]. In contrast, proponents of the categorical approach view emotions as innate, evolutionary predetermined mechanisms indispensable for the survival of the organism. The evidence for the existence of discrete emotions is multi-faceted. Keltner et al. [125] provide evidence indicating that facial expressions are perceived categorically and linked to distinct brain regions, autonomic activity, and evoked responses in others. Izard et al. [121] describe the adaptive functions of discrete emotions that provide evidence for usefulness of the approach. Finally, Panksepp [120] postulates there are basic emotional systems in the brain that produce emotions as natural kinds; these systems are: seeking/expectancy (perhaps *Interest* emotion); rage/anger; fear/anxiety; panic/separation (possibly related to *Sadness*), play/joy, care/nurturance, and lust/sexuality.

Although these two views approach emotion differently, they are not necessarily contradictory [123, 125, 126]. For our purposes, we find the categorical approach to emotions more suitable, as it allows modeling each emotion as a mechanism with its own adaptive function, antecedents, and a set of responses. However, on occasion, the dimensional view may prove useful as well, e.g., when establishing the influence of personality on emotions: in particular, neuroticism has been linked to a propensity for experiencing negative emotions, and extraversion – positive.

3.3.1.2 Functions of Emotions

There is a certain number of functions characteristic of emotions as a whole, and there are those that are more or less specific to distinct emotions. Although the list of common functions of emotions can be very long, as it would include those related to organizing, motivating and managing physiological, expressive and cognitive processes, as well as generating and coping with subjective feelings, we will restrict the one in this section to those functions that are relevant for autonomous behavior-based systems, namely those that relate to perception and action.

1. **Coordination of attention and perception.** Emotions generally arise in response to a situation relevant for survival, and serve to focus the attention of an individual to such a situation, making it salient, and allow examining it more closely (figuratively and literally), for a better choice of action. For example, fear narrows the focus of attention to potentially dangerous objects and makes the individual more sensitive to perceiving disturbing cues; interest provides a mechanism of selective attention, which keeps attention from straying more or less randomly through a wide variety of stimuli [121].

2. **Motivating and guiding action.** This function is implicitly present in most definitions of emotion, as many of them include a behavioral response to an environmental contingency. Emotions also allow flexibility of behavioral responses to stimuli – they provide a simple interface between sensory inputs and action systems [127]. As we strive to improve robotic behavior by providing a flexible link between perception and action, this function is of primary interest.

3. **Expressive and communicative function.** The external display of emotions via facial expressions, voice and body posture may be used to read an individual's emotional state and therefore anticipate imminent behavior, better assess the current situation, or respond to an individual's needs [126]. E.g., sadness communicates a need for help, fear – an external, situational threat, and anger – a potential threat from the one expressing it. Social bonding may be viewed as an extension of this function, where, e.g., sadness and empathy would promote bonding at the time of grief, and joy – at the time of success.

Functions that are specific to certain emotions can provide us with an understanding of what influence an emotion would make on the current state of an autonomous robot, given an appropriate stimulus. **Table 7** lists a number of emotions, their functions, and in addition to functions, a core relational theme which provides a summarized, overall meaning or cause for most of these emotions, as described by Lazarus [128], with the exception of Interest for which no core relational theme was provided.

Table 7: Functions and Core Relational Themes for Select Emotions

Emotion	Function	Core Relational Theme
Interest	Motivates exploration and learning; guarantees engagement in the environment; provides a mechanism of selected attention; motivates approach responses [121]	Not available
Joy	Heightens an openness to experience; contributes to affilliative behaviors and strengthening of social bonds; smile signals friendly interaction [121]; also invites others to participate [129]	Making reasonable progress towards realization of a goal
Sadness	Strengthens social bonds; slows cognitive and other processes that allows to look for the source of the problem and perhaps avoid it in the future; signals a need for help [121]	Having experienced an irrevocable loss
Anger	Mobilizes and sustains energy (resources) [121]; organizes and regulates processes related to self-defense and mastery; regulates social and interpersonal behaviors [130]	A demeaning offense against me and mine
Fear	Promotes escape/avoidance behavior; organizes and directs perceptual and cognitive processes [121]	An immediate, concrete, and overwhelming physical danger (Fright); facing uncertain, existential threat (Anxiety)
Disgust	Promotes distancing from the offending, potentially contaminating, stimuli [131]	Taking in or being too close to an indigestible object or idea

3.3.1.3 Primary vs. Secondary Emotions

Emotions also differ in the way they are generated. A subdivision widely accepted among emotion theorists is that between primary and secondary emotions. Although the exact emotions that belong to each category, as well as the number of categories itself differ between various researchers, the main idea remains the same: emotions can be separated into fast, automatic, hard-wired reactions and slower, cognitive and more flexible ones. For example, Panksepp [120] differentiates between “the Blue-Ribbon, Grade-A Emotions” and “the Higher Sentiments”. The “Grade-A emotions” have evolutionarily evolved earlier and are conceptualized as sensorimotor emotional command circuits in the brain (they include emotions like fear, anger, sadness, joy, and interest). The Higher Sentiments, such as shame, guilt, pride, gratitude and others, are more subtle social emotions that result from a more recent evolutionary expansion of the brain. LeDoux [17] distinguishes between “low road”/”fast pathway” of emotion generation that bypasses the cortex and results in “quick and dirty” processing and “high road” that benefits from cortical processing that is more detailed and accurate, albeit slower.

For this work, we will be mostly concerned with primary emotions, as they present a more straightforward link to behavior, and, providing faster responses, are more suitable for the robotics domain. However, as this research progresses, we may consider including “more subtle social emotions”, especially for the benefit of human-robot interaction domain.

3.3.1.4 Influence by Traits

Personality has a profound influence on emotions. Lazarus [30] says that an emotion is a result of not just a property of the environment, but depends on the conjunction of the situation and personality. There are a number of ways where the personality influence comes into play. First of all, people with different personalities may find different events more salient than others, thus generating emotions in response to different events. Some individuals may view situations in a more negative or positive light, and therefore be more prone to experiencing more negative or positive emotions, respectively. Another way for personality to influence emotions is in lowering thresholds for response patterns characterizing a particular emotion [3]. For example, those who score high on Anxiety (a facet of Neuroticism) are more prone to experiencing and exhibiting fear than low-scorers. Likewise, the Positive Emotions facet of Extraversion dimension reflects a propensity of an individual to experience positive emotions. Response patterns often include action tendencies – impulses to engage in a certain behavior; personality may govern both the type of tendencies brought about by an emotion, and the extent to which these tendencies will result in a particular behavior. Personality also contributes to coherence, consistency and predictability in emotional reactions and responses [132].

Finally, Davidson [133] proposes that the following variables, influenced by personality, describe "affective chronometry" – the temporal dynamics of emotional responses:

1. Threshold of eliciting stimuli for a particular emotion (an activation point);
2. Peak or amplitude of the response;
3. Rise time to peak (how long it takes the emotions to reach its peak);
4. Recovery time (decay rate);

In this research, we will provide the means to account for personality differences in the first three of the aforementioned variables.

3.3.1.5 Influence by Moods

Although no systematic data exists, the supposition is that moods may influence emotions by: 1) increasing the probability that similarly valenced emotions may be triggered [10]; 2) lowering the thresholds for experiencing a particular emotion [3]. For this research, the moods will increase the sensitivity of certain emotions to emotion-eliciting stimuli.

3.3.1.6 Influence by Attitudes

As far as emotions are concerned, according to [3], they stand in a "close and reciprocal relationship" with attitudes, where a sentiment may originate in an comparable emotion, and emotions can be brought about by meeting or thinking about the object of a sentiment. In a sense, affective attitudes carry with them latent emotions that can be brought to life when their objects become salient. For this research, an attitude may invoke a corresponding emotion by adding emotion-eliciting characteristics to a stimulus; in particular, attitude strength will be one of the components used in calculating the overall stimulus strength.

3.3.1.7 Implications for TAME

The aforementioned psychological findings allow us to make the following design decisions:

1. The emotion component will be modeled as a set of distinct emotions, each having a different function and different effects on action, and expressive behavior (in the case of human-robot interaction).

2. The primary emotions to be modeled are: *Anger, Fear, Disgust, Sadness, Interest and Joy*. In terms of their interaction with other affective components, *Anger, Fear, Disgust* and *Sadness* will be regarded as negative, and *Joy* and

Interest as positive. In accordance with the “fast pathway” view of emotion generation, no cognitive reasoning will be involved in emotion generation.

3. Being short-term and object-specific, the emotions will be invoked in response to environmental stimuli, and will last either until the stimulus disappears (with a short-term lingering effect), or until they decay in time, whichever is faster. The inputs to this component are: stimuli characteristics (sensor data or processed stimuli information) and the values of traits, moods and attitudes.
4. Personality, mood and attitude influences will be taken into account. In particular, traits will affect sensitivity to stimuli (activation point), amplitude, and time rise to peak; moods will only influence sensitivity to stimuli for similarly valenced emotions (negative mood will affect negative emotions, and positive mood will affect positive ones); finally, attitude will play a part in determining the stimulus strength and the type of emotion generated.

3.3.1.8 Application to HRI

Emotions have an immense potential in the field of human-robot interaction, as evidenced by the number of robotic systems which include emotions in one way or another. First, emotions have a high expressive value, providing interacting humans with cues not only regarding the internal state of the robot, but also regarding the situation they are in. Second, emotions play a crucial role in interpersonal relationships, promoting a sense of camaraderie, rapport and acceptance. Finally, emotional expressiveness can help make robots less alien, more understandable, and contribute to creating a greater illusion of life. **Table 8** provides examples of how each of the emotions can be utilized in the human-robot interaction domain, along with the thematic roles it can play.

Table 8: Examples of applying emotions to Human-Robot Interaction

Emotion	Example	Role
Joy	If the robot rejoices in events others find happy, it could promote the team spirit, and facilitate the robot's acceptance as a part of the team.	*Co-worker* *Co-inhabitant*
Interest	May help attract people's attention to certain points of interest otherwise not immediately salient to humans; expressing interest in the lives of those around it on a daily basis would help increase acceptance and attachment.	*Co-explorer* *Co-inhabitant*
Fear	May signal imminent danger to nearby people, and be more persuasive than words alone, in case, for instance, an evacuation is required.	*Co-protector*
Disgust	May alert a human to the presence of some noxious stimulus, which, though not necessarily hazardous, would still be best avoided.	*Co-protector* *Co-worker* *Co-inhabitant*
Anger	By showing a willingness to protect "its own", the robot may alleviate some fear its "charges" may experience in a dangerous situation.	*Co-protector*
Sadness	Similarly to joy, may promote a sense of belonging to a group, through empathizing with a goal failure, or with problems a team/family member is experiencing.	*Co-worker* *Co-inhabitant*

3.3.2 REPRESENTATION AND GENERATION

Emotion values are not user-defined, but are dynamically generated throughout a robot's mission based on the presence and strength of environmental stimuli. Picard [20] identifies a number of properties of emotions that would be desirable to model in affective systems. The following subset of them will be modeled in *TAME*:

1. A property of activation, which refers to certain stimulus strength below which the emotional circuits are not activated.

2. A property of saturation, which refers to the upper bound of an emotion, after which, regardless of the increasing stimulus strength, the emotion doesn't rise

any more. This upper bound can be defined by physical capabilities (e.g., the heart can beat only so fast in response to an emotion).

3. A property of response decay, which states that emotions decay naturally over time unless they are re-stimulated.

4. A property of linearity. Picard [20] suggests that emotions can be modeled as linear under certain conditions; due to the properties of activation and saturation, the emotions will approximate linearity only for certain stimulus strength range, and will approach a sigmoid at the edges of this range. Picard proposes a sigmoidal non-linearity [20] to account for a number of these properties.

Each emotion's intensities are stored in the emotion intensity matrix:

$$\vec{E} = [E_i]$$

where $0 \le E_i \le g_i$, the value E_i represents the intensity of a currently active emotion, with 0 being the absence of emotion, and g_i is the experimentally defined upper bound for emotion *i*.

The basic emotions to be modeled in *TAME* are *Fear, Disgust, Anger, Sadness, Joy* and *Interest*, therefore:

$$\vec{E} = \begin{bmatrix} E_{\text{Fear}} \\ E_{\text{Disgust}} \\ E_{\text{Anger}} \\ E_{\text{Sadness}} \\ E_{\text{Joy}} \\ E_{\text{Interest}} \end{bmatrix}$$

3.3.2.1 Eliciting Environmental Stimuli

The exact nature of environmental stimuli will depend on a particular environment; however, we can identify certain general features that can be attributed to a majority of stimuli that an autonomous robot may encounter. These features will be used to identify the type of emotion generated and its intensity. First, **type of stimulus** refers to the intrinsic quality of a stimulus to invoke a particular emotion. It could be determined by a certain property that the stimulus possesses, e.g., color or shape, or a combination of qualities (e.g., a receding likable object may lead to sadness). Additionally, the result of a cognitive assessment of a situation can also serve as a stimulus type – for example, a successful accomplishment of a goal can be interpreted as a joyful stimulus. As a simplifying assumption, type signifies that an emotion-eliciting object is present, but does not determine the intensity of emotion. The same stimulus may cause multiple emotions; for example, anger and fear often appear in response to the same object.

Other stimulus features, such as the size of stimulus, its brightness, shape, distance to it, and others determine the intensity of the generated emotion. According to McFarland et al. [134], certain stimulus features add a specific contribution independent of other features; such features are additive in their effect. These additive heterogeneous features can be combined into a single index, called "cue strength", and represented along a single dimension. Other independent features may be represented along a different dimension (e.g., distance), and can be combined multiplicatively with the cue strength. Although this characterization of stimulus properties was originally applied to motivational space, it seems reasonable and applicable in emotion generation as well. The more permanent object attributes, e.g., size, category, shape, etc. also serve as determinants for forming long-lasting attitudes towards these objects (described in detail in subsection 3.5.2, *Representation and Formation*).

For the robotics domain, the following cues (i.e., dimensions) can be identified:

1. **Physical Properties Cue**. This category will refer to physical properties of a stimulus. Depending on different robotics platforms and sensors, it can include a number of features:

(a) Object size – can be determined as relative to robot size;

(b) Brightness/hue (e.g., if a certain color category serves as an elicitor of a particular emotion, then the extent to which the object color falls within this category can be a feature included into the overall physical properties cue);

(c) Loudness (for auditory sensors);

(d) Shape and others, as applicable, depending on the environment and physical platform.

2. **Position Cue**. This category refers to the relative position and orientation of the eliciting stimulus to the robot and may include the following:

   a) Distance to the stimulus – the upper bound can be determined by sensor range; the closer the stimulus, the greater the cue;

   b) Velocity of stimulus approach and withdrawal, and its acceleration/deceleration (in case of a moving stimulus);

   c) Stimulus orientation (e.g., if there's a clear "front" and "back" of the eliciting stimulus, then the object with its "back" to the robot may not present as much danger as the one oriented towards the robot);

   d) Others, as applicable, depending on the environment and physical platform.

3. **Attitudinal Cue**. If the robot had developed an attitude towards a particular object, then the strength of this attitude can be used in calculating the overall stimulus strength. These attitudes towards particular objects can be based on certain more or less permanent object attributes, and the resulting attitude strength for a particular emotion serves as attitudinal cue. E.g., if a robot "likes" an object/person, then the resulting overall stimulus strength will be greater when generating positive emotions, than without this attitude, all other cues being equal. Attitude can also serve as the only cue involved in emotion generation.

4. **Situational Cue**. A cognitive assessment of a situation can serve as a cue, even if a physical object is not present. For example, a successful accomplishment of a goal would illicit *Joy*, and withdrawal of a "loved" object would cause *Sadness*. The strength of this cue can be determined by goal importance, degree of success, and other intrinsic situational properties.

As a simplifying assumption, only one stimulus at a time is considered for a particular emotion, therefore at any point in time the stimuli strength is stored in the stimulus strength matrix, $\vec{s} = [s_i]$, where $0 \le s_i < \infty$, s_i is the stimulus strength for emotion *I*; 0 signifies absence of a stimulus. The stimulus strength for emotion *i* is then defined as a product of all available stimulus cues, each of which is an additive combination of stimulus features comprising a particular cue:

$$s_i = \prod_{c=1}^{m} \left(\sum_{f=1}^{n} feature_{fc} * k_{fc} \right) \qquad (5)$$

where s_i is stimulus strength for emotion *i*, *m* is the number of cues for the stimulus in question, *n* is the number of features in cue *c*, and *k* is a scaling factor to bring each feature to the same units. Although in theory the total stimulus strength s_i could go to infinity, in practice it is likely that the features will have certain lower and upper bounds (e.g., the sensors can only reach so far in identifying the distance to the object, or to discern an object of a very small size). These bounds would be empirically determined and will depend on both the platform and the environment.

3.3.2.1.1 Stimuli Strength Illustration

Suppose a large robot presents a source of threat to a robot augmented with emotional capabilities. The affective robot's available sensor data allows to determine the threatening robot's size and loudness, comprising together a Physical Properties Cue, and the distance to it, which represents a Position Cue. To illustrate the combination of these cue strengths into the total stimulus strength, several isoclines for total stimulus strengths were plotted in **Figure 7.**

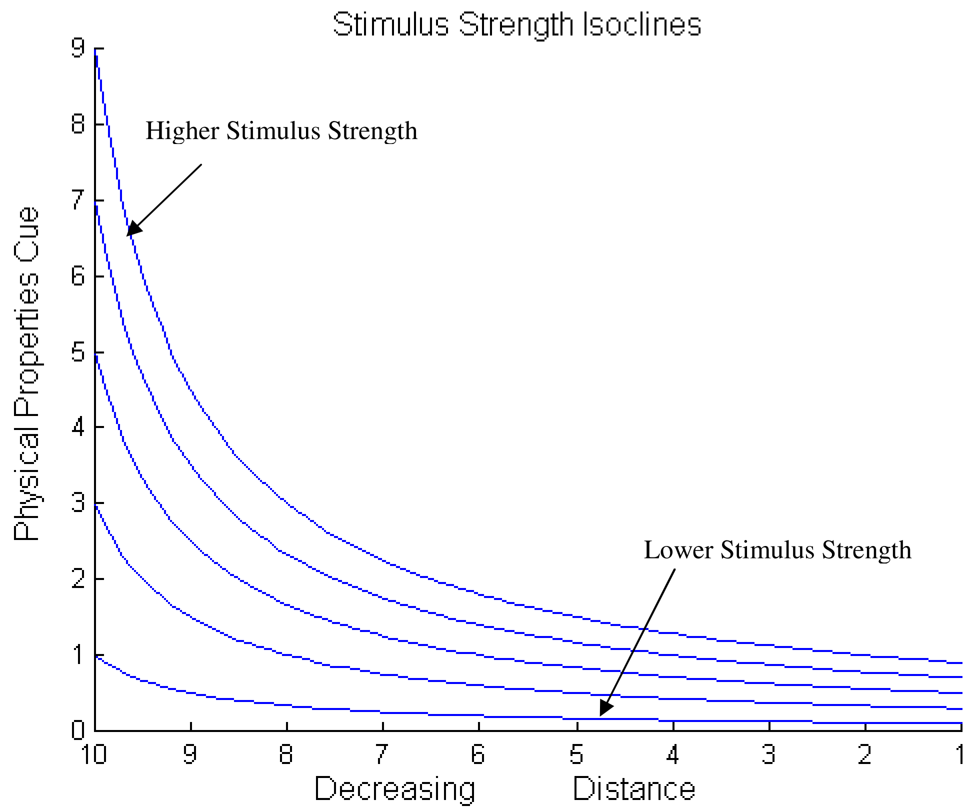


Figure 7: Stimulus Strength Isoclines Based on Distance and Physical Properties Cues.

In this figure, physical Properties Cue is plotted along the Y axis; decreasing distance is used as the only feature of Position Cue for clarity and is plotted along the X axis. The isoclines correspond to different relative stimulus strengths. As the stimulus strength is the primary determinant of emotion, this figure also illustrates what relative effect the multiplicative combination of cue strengths will have on emotion generation. In particular, the graph shows that as the distance grows increasingly smaller, only a low intensity physical properties cue (e.g., a small size of an object) would be needed to produce an equivalent stimulus strength (and consequently, emotion), and vice versa, at a great distance, objects with greater physical properties cue will be required to reach the same stimulus strength. For certain emotions and tasks, an inverse relationship between some of the cues and stimulus strengths is also true; for example, it is possible that for some tasks a smaller object would be of more interest than a large one (Physical Properties Cue).

An alternative mechanism for calculating the overall stimulus strength can be a weighted average of all present cues (normalized). Although not biologically inspired, this method is simpler to implement, and has an advantage of taking relative importance of cues into account; for example, object orientation may be more important for generation of *Fear*, but not of *Disgust*. This latter approach was taken during the software design and implementation stage.

3.3.2.2 Emotion Generation

Now that the strength of the eliciting stimulus is known, the base emotion can be calculated. The desired function for base emotion generation resembles a sigmoid curve. This function supports the aforementioned properties of emotions and the influences by traits and moods in the following way:

1. **The property of activation**. The function has a lower bound (that corresponds to the absence of emotion), and at very small inputs, the output will be close to the lower bound until a certain activation point, after which it will grow slowly at first, but approximate linear closer to the mid-point of the curve.

2. **The property of linearity.** In the middle section of the curve (for average inputs) the output will be approximately linear.

3. **The property of saturation.** As the input stimulus grows to infinity, the resulting emotional value should approach a finite upper bound.

4. **Personality and Mood Influences**. This function also incorporates influences from other affective components by allowing variations in the amplitude, maximum slope and activation point.

Picard [20] proposes an inverse exponential function to account for these properties. Although inspired by it, we chose a combination of two exponential functions as the desired function, as it can accommodate the influences due to personality and moods in a way that fits our overall framework better. The following equation calculates the base emotion value, based on the current stimulus strength and mood and personality influences:

$$E_{i,base} = \begin{cases} e^{(s_i - a_i)/d} - e^{(-a_i)/d_i}, & if \quad a_i \le s_i < (a_i + b_i)/2 \\ g_i - e^{(s_i - b_i)/d_i}, & if \quad s_i \ge (a_i + b_i)/2 \end{cases}$$
$$where \quad b_i = 2 \cdot d_i \cdot \ln(g_i + e^{(-a_i)/d_i})/2 + a_i \tag{6}$$

where $E_{i,base}$ is the base emotion value for emotion *i*, s_i is the strength of stimulus eliciting emotion *i*, a_i is the variable that controls the activation point for emotion *i*, d_i is the variable that controls the maximum slope for emotion *i*, *g* is the amplitude of emotion *i,* and b_i is the break-point, at which the emotion reverses its rate of growth. As variables a, d and g can change with personality and mood influences, they will be referred to as "*emotion generation variables*" from now on.

Figure 8 presents a general shape of the resulting curve. When the stimulus strength is very low, no emotion is generated; then, as the stimulus strength increases, the emotion starts rising slowly, becomes approximately linear in the middle where the rise is faster; finally, as the stimulus strength gets larger, the growth of emotion slows down, until the emotion value reaches its peak. After that, further increases in stimulus strength will not produce any additional increase in the emotion.

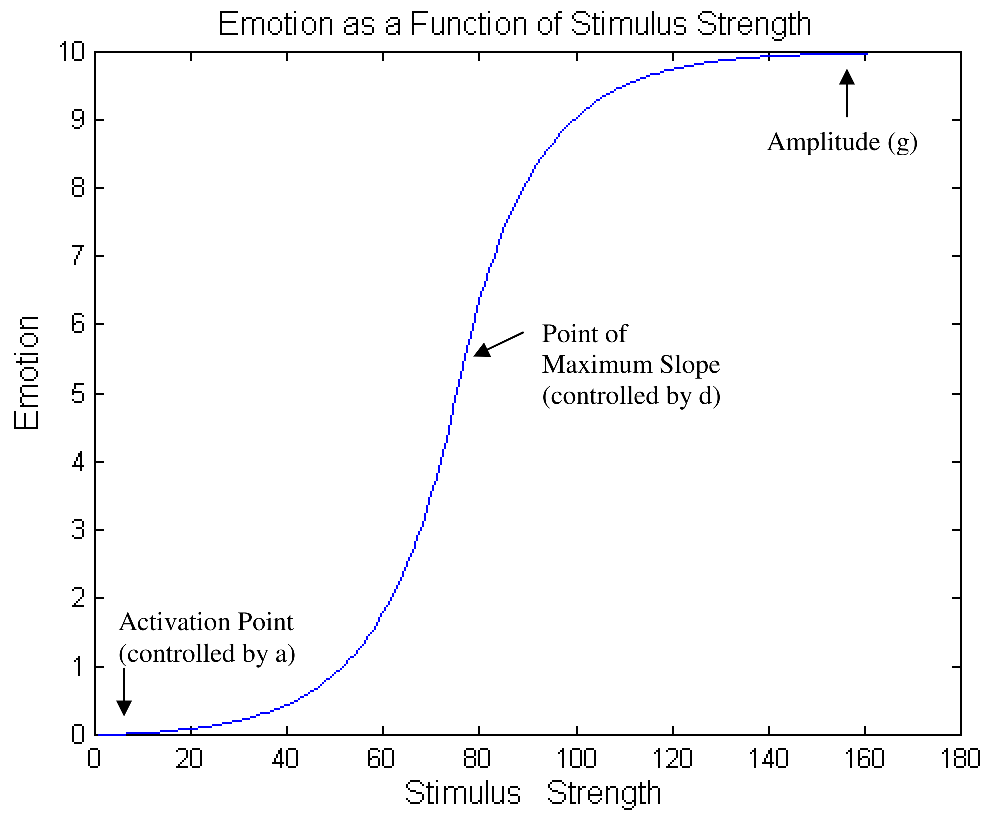


Figure 8: Emotion Generation Based on Stimulus Strength.

Figure 9 presents an illustration of how emotion is generated for different stimuli strengths at a single point in time, as the emotion generation variables (amplitude, activation, and slope) are held constant (neutral). In this figure, four different stimuli appear and disappear as time progresses; their proximity in time is for illustration purposes only, and does not reflect the frequency with which emotions are normally generated. Scaled values of stimuli strength are plotted in blue dotted line, and the generated emotion values are plotted in red solid line. We can see that at low stimulus strength, the generated emotion grows slowly, and is significantly lower that the corresponding stimulus strength. As the stimulus strength is closer to the middle of the curve, the generated emotion grows significantly faster. Finally, as the emotion approaches its saturation point with higher stimulus strength, the growth slows down again until the emotion reaches its saturation.

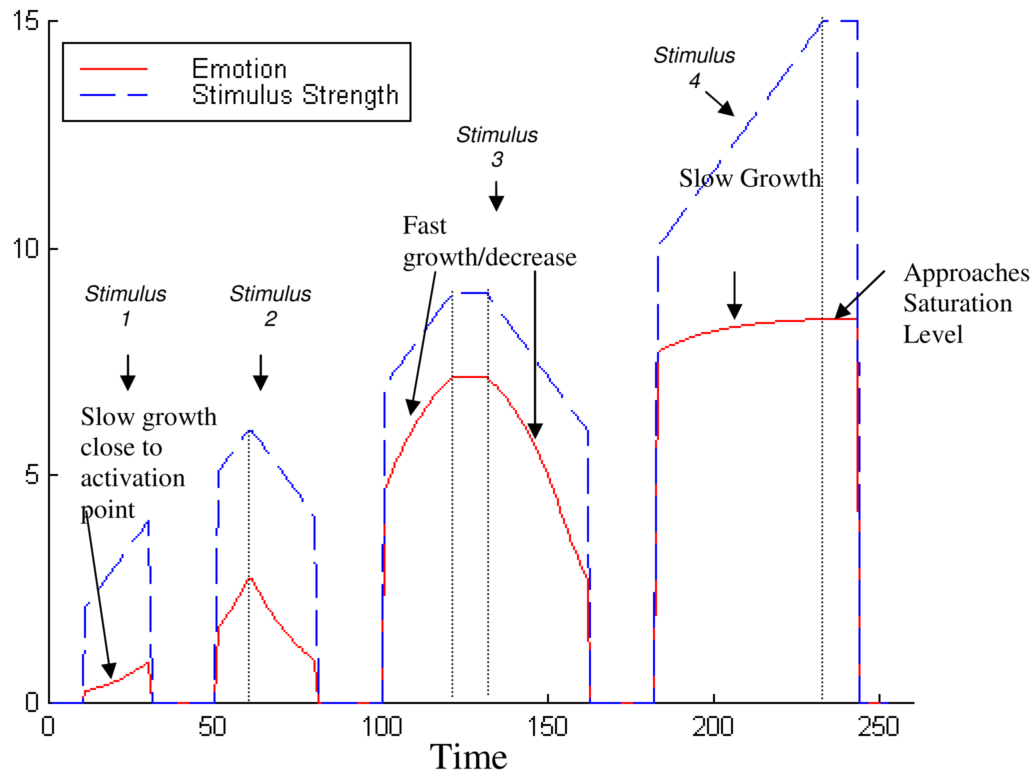


Figure 9: Effect of different strength stimuli on emotion generation.

3.3.2.2.1 Personality Influence on Base Emotion Generation

Personality may affect each of the emotion generation variables. The peak, or amplitude, of an emotion can be varied by changing the upper bound (*g*). The sensitivity to stimulus can be varied by changing the activation point (controlled by *a*). This variable will vary the point at which the emotion will start to rise; e.g., for individuals high in neuroticism, even smaller stimulus strengths will produce a certain level of emotion, whereas in those with a lower value of neuroticism, the activation point at which a stimulus will start generating emotion generation will be higher. Finally, the maximum slope of the curve can be varied to affect the "rise time to peak", controlling the rate at which an emotion rises with increasing stimulus strength. Although the function is not time-based, due to the fact that the stimulus strength is expected to vary with time (e.g., distance will be increasing or decreasing), varying the slope will also vary how long it will take for the emotion to rise to a certain level.

For the present, a linear mapping from personality traits to each of these variables will be used, and can be reevaluated later. In order to obtain such a mapping, two sets of data points are required. The upper and lower bounds for each trait have been established earlier, in section 3.2.2; the upper and lower bounds for the emotion generation variables will need to be determined experimentally, based on a variety of robotic platforms and potential stimuli. Given these two sets of data points, we can fit them into a line of a general form of y=ax+b, thus providing a mapping from the trait values onto the emotion generation variables. Therefore, each of the traits can be mapped to each of the variables through linear functions that produce a line: $f_{i,g}(p_j)$ for the amplitude, $f_{i,d}(p_j)$ for the maximum slope, and $f_{i,a}(p_j)$ for the activation point.

Although very little systematic research exists on how specific individual differences influence emotion generation [29], Watson et al. [4] provide evidence for more general correlations between the Big Five personality dimensions and Positive and Negative Affect. For example, both Neuroticism and Agreeableness may influence the generation of fear, where the trait of Neuroticism produces a direct influence (has a strong positive correlation with Negative Affect), and the trait of Agreeableness produces an inverse influence (has a negative correlation with Negative Affect). To allow greater flexibility, each trait can have a direct, inverse, or no influence on each emotion generation variable (activation point a, maximum slope d, and amplitude g). For clarity, three personality-emotion dependency matrices are defined, one for each variable:

$$\overrightarrow{pg} = [pg_{ij}] \text{, where } pg_{ij} \in \{-1,0,1\};$$

$$\overrightarrow{pd} = [pd_{ij}] \text{, where } pd_{ij} \in \{-1,0,1\};$$

$$\overrightarrow{pa} = [pa_{ij}] \text{, where } pa_{ij} \in \{-1,0,1\};$$

In each of these matrices, 0 specifies an absence of influence of trait *j* on emotion generation variables (g, d, and a, correspondingly) for emotion i, +1 refers to direct influence, and –1 to inverse influence. For example, following our prior example of Fear generation, $pg_{Fear,\ Neuroticism}$ = 1 (direct influence of Neuroticism on fear amplitude), $pg_{Fear,\ Agreeableness}$ = -1 (inverse influence of Agreeableness on fear amplitude), and $pg_{Fear,\ Openness}$, $pg_{Fear,\ Extraversion}$, and $pg_{Fear,\ Conscientiousness}$ = 0 (no influence). This representation allows for a greater flexibility in specifying the relationship between personality and emotion generation, useful for when either psychological data become available, or for adaption to different robotic tasks and environments. In case of activation point a, however, we would expect an opposite relationship: the activation point would be lower with higher values of Neuroticism (thus increasing the sensitivity to fear-invoking stimuli), and with lower values of Agreeableness (lowering the sensitivity).

Taking into consideration personality/emotion dependency matrices, the linear relationship between personality traits and emotion generation variables g (amplitude), d (slope) and a (activation variable) can be expressed by the following functions:

$$f_{i,g}(p_j) = \begin{cases} \left(\dfrac{g_{i,upper} - g_{i,lower}}{p_{j,upper} - p_{j,lower}}\right) * \left(p_j - p_{j,lower}\right) + g_{i,upper}, & \text{if } pg_{ij} = 1 \\ 0, & \text{if } pg_{ij} = 0 \\ \left(\dfrac{g_{i,lower} - g_{i,upper}}{p_{j,upper} - p_{j,lower}}\right) * \left(p_j - p_{j,lower}\right) + g_{i,upper}, & \text{if } pg_{ij} = -1 \end{cases} \quad (7)$$

where $f_{i,g}(p_j)$ is the linear mapping function relating personality trait *j* to amplitude *g* for emotion *i*, $g_{i,upper}$ and $g_{i,lower}$ are upper and lower bounds for amplitude for generating emotion *i*, $p_{j,upper}$ and $p_{j,lower}$ are upper and lower bounds for trait *j*, p_j is the intensity value for trait *j*, and $\overrightarrow{pg}$ is the personality-emotion amplitude dependency matrix.

$$f_{i,d}(p_j)=\begin{cases}\left(\dfrac{d_{i,upper}-d_{i,lower}}{p_{j,upper}-p_{j,lower}}\right)*\left(p_j-p_{j,lower}\right)+d_{i,upper}, & \text{if } pd_{ij}=1\\ 0, & \text{if } pd_{ij}=0\\ \left(\dfrac{d_{i,lower}-d_{i,upper}}{p_{j,upper}-p_{j,lower}}\right)*\left(p_j-p_{j,lower}\right)+d_{i,upper}, & \text{if } pd_{ij}=-1\end{cases} \tag{8}$$

where $f_{i,d}(p_j)$ is the linear mapping function relating personality trait *j* to slope *d* for emotion *i*, $d_{i,upper}$ and $d_{i,lower}$ are upper and lower bounds for slope for generating emotion *i*, $p_{j,upper}$ and $p_{j,lower}$ are upper and lower bounds for trait *j*, p_j is the intensity value for trait *j,* and $\overrightarrow{pd}$ is the personality-emotion slope dependency matrix.

$$f_{i,a}(p_j)=\begin{cases}\left(\dfrac{a_{i,upper}-a_{i,lower}}{p_{j,upper}-p_{j,lower}}\right)*\left(p_j-p_{j,lower}\right)+a_{i,upper}, & \text{if } pa_{ij}=1\\ 0, & \text{if } pa_{ij}=0\\ \left(\dfrac{a_{i,lower}-a_{i,upper}}{p_{j,upper}-p_{j,lower}}\right)*\left(p_j-p_{j,lower}\right)+a_{i,upper}, & \text{if } pa_{ij}=-1\end{cases} \tag{9}$$

where $f_{i,a}(p_j)$ is the linear mapping function relating personality trait *j* to activation point *a* for emotion *i*, $a_{i,upper}$ and $a_{i,lower}$ are upper and lower bounds for activation point for generating emotion *i*, $p_{j,upper}$ and $p_{j,lower}$ are upper and lower bounds for trait *j*, p_j is the intensity value for trait *j*, and $\overrightarrow{pa}$ is the personality-emotion activation point dependency matrix.

Continuing with the example of Fear generation, let's suppose Neuroticism (*N*) produces a direct effect on Fear amplitude; Neuroticism upper and lower bounds are within +/- 2 SD of mean ($p_{N,upper}$ = 79.1+2*21.2 = 121.5 and $p_{N,lower}$ = 79.1-2*21.2 = 36.7), Fear amplitude upper bound $g_{Fear,upper}$=10 and lower bound $g_{Fear,lower}$ = 6. Then, if the robot's Neuroticism value p_N is 94, the value for amplitude $g_{Fear,N}$ is:

$$g_{Fear,N} = \left(\frac{10-6}{121.5-36.7}\right) * (94-36.7) + 6 = 8.7$$

The trait of Agreeableness (*A*), on the other hand, has an inverse effect on Fear amplitude. Supposing Agreeableness upper and lower bounds are within +/- 2 SD of mean ($p_{A,upper}$ = 124.3+2*15.8 = 155.6 and $p_{A,lower}$ = 124.3-2*15.8 = 92.4), Fear amplitude upper bound $g_{Fear,upper}$ =10 and lower bound $g_{Fear,lower}$ = 6, and Agreeableness current value p_A is 100 (low), then the resultant value for amplitude $g_{Fear,A}$ is:

$$g_{Fear,A} = \left(\frac{6-10}{121.5-36.7}\right) * (100-92.4) + 6 = 9.5$$

Figure 10 shows the difference in the shape of the base emotion generation function as the influencing personality trait goes from the lowest to the highest. In this figure, a trait has a direct influence on emotion generation (e.g., Neuroticism's influence on Fear) and the value of mood is kept constant. As the value of the influencing trait goes up, the amplitude of the emotion gets bigger, the slope is steeper, and the activation point is lower (the sensitivity to the stimulus is higher). For example, if the trait under consideration is Neuroticism, and the resulting emotion is fear, then the individuals with the low value of Neuroticism will not be as sensitive to fear-inducing stimuli (it will take a stronger stimulus for them to start experiencing emotion), the resulting peak emotion will not be as high, and it won't rise quite as fast with increasing stimulus strength as that of their High Neuroticism counterparts.

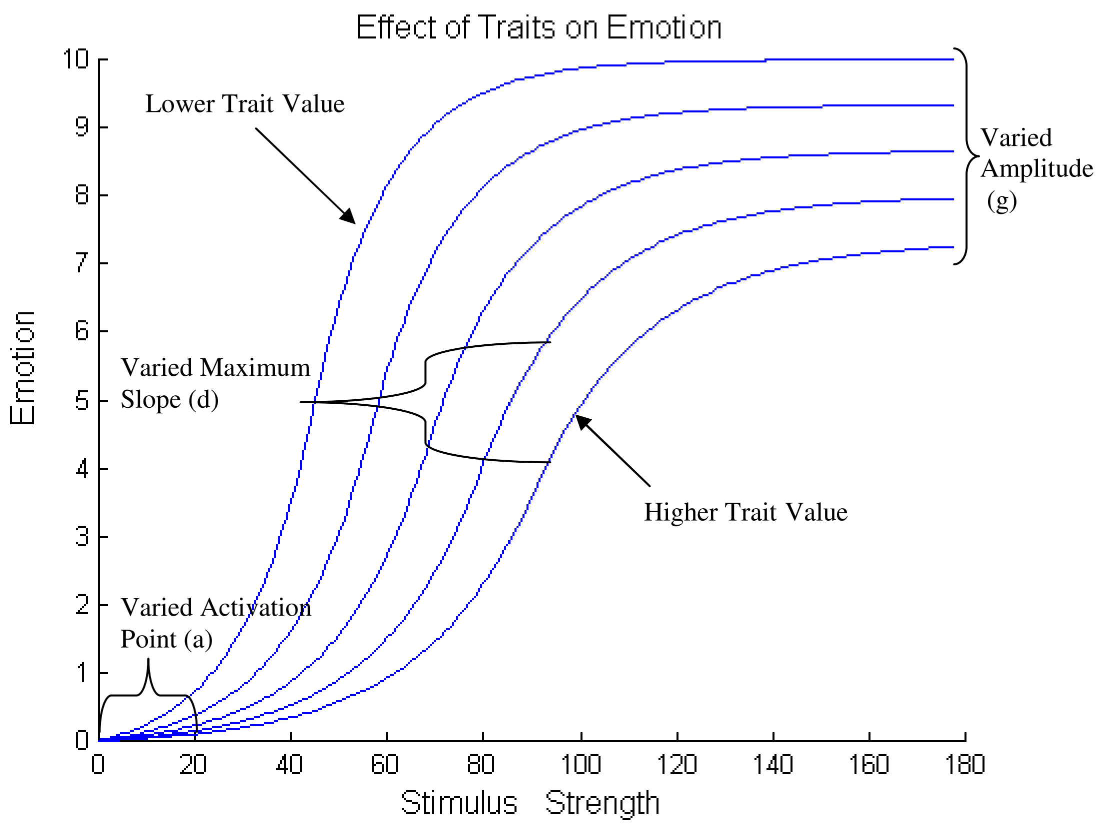


Figure 10: Effect of Personality on Emotion Generation in terms of emotion amplitude, activation point and slope.

It is possible that multiple traits may affect the same emotion generation variable. Then the overall personality influence on an emotion generation variable is defined as the average emotion generation variable of all personality traits influencing that emotion (given the linear mapping functions $f_{i,g}(p_j)$, $f_{i,d}(p_j)$, and $f_{i,a}(p_j)$ that map personality values to emotion generation variables g, d and a, respectively):

$$g_i = \frac{1}{N}\sum_{j=1}^{N} f_{i,j}(p_j) \qquad (10)$$

where g_i is the amplitude for emotion i, $f_{i,g}(p_j)$ is the function that provides a linear mapping between trait j and the emotion generation variable g, N is the total number of traits, i.e. 5, in the trait intensity matrix $\vec{p}$, and p_j is the intensity value for trait j.

$$d_i = \frac{1}{N}\sum_{j=1}^{N} f_{i,j}(p_j) \qquad (11)$$

where d_i is the maximum slope variable for emotion *i*, $f_{i,d}(p_j)$ is the function that provides a linear mapping between trait *i* and the emotion generation variable d, *N* is the total number of traits, i.e. 5, in the trait intensity matrix $\vec{p}$, and p_j is the intensity value for trait *j*.

$$a_{i,trait} = \frac{1}{N}\sum_{j=1}^{N} f_{i,j}(p_j) \qquad (12)$$

where $a_{i,\ trait}$ is the activation point variable for emotion *i*, $f_{i,a}$ (p_j) is the function that provides a linear mapping between trait *j* and the emotion generation variable a, *N* is the total number of traits, i.e. 5, in the trait intensity matrix $\vec{p}$, and p_j is the intensity value for trait *j*. The activation point variable *a* is influenced both by personality and moods, and $a_{i,trait}$ presents the trait-based portion of this variable.

Finally, to conclude our example of calculating trait effect on Fear amplitude, suppose Neuroticism and Agreeableness are the only traits that exert an influence on Fear amplitude. Then the overall trait-based Fear amplitude g_{Fear} is the average of $g_{Fear,A}$ and $g_{Fear,A}$ = (8.7+9.5)/2 = 9.1.

3.3.2.2.2 Mood Influence on Base Emotion Generation

According to Frijda [3], moods may influence emotions by lowering the threshold for their elicitation. Therefore, for the present, mood intensity will affect sensitivity to eliciting stimuli by changing the activation point for a corresponding emotion. In particular, the negative mood will make the robot more sensitive to negatively valenced stimuli (those eliciting negative emotions), and the positive mood will make the robot more sensitive to positively valenced stimuli. Other potential effects may also be explored later, should further readings or experimentation prove them useful.

The mood-based activation variable $a_{i,mood}$ can be calculated in a similar manner to personality influence on emotion generation variables. In particular, given a mood intensity m_j (see section 3.4.2, Representation and Generation of Moods), the desired upper and lower bounds for mood intensity and the desired upper and lower bounds for the activation point variable, $a_{i,mood}$ can be produced by linear mapping from the corresponding mood intensity. In order to flexibly specify mood influence on activation point variable (a), a mood/activation point matrix reflecting correlation between emotion *i* and mood *j* is defined:

$$\overrightarrow{ma} = [ma_{ij}] \text{, where } ma_{ij} \in \{-1,0,1\} \text{;}$$

In this matrix, 1 stands for direct influence, -1 for inverse influence, and 0 for absence of influence. Suppose negative mood has an inverse impact on activation point used for generating negative emotions (Fear, Anger, Disgust and Sadness): the higher the Negative Mood, the smaller stimulus strength would be required to initiate generation of negative emotions. Similarly, positive mood would affect positive emotions (Joy and Interest) in the same manner. The mood/activation point dependency matrix reflecting this relationship is given in **Table 9**.

Table 9: A sample mood/activation point dependency matrix

	Fear	**Anger**	**Sadness**	**Disgust**	**Joy**	**Interest**
Positive Mood	0	0	0	0	-1	-1
Negative Mood	-1	-1	-1	-1	0	0

The function mapping mood to activation point is similar to the one providing linear mapping between traits and activation point (subsection 3.3.2.2.1, *Personality Influence on Base Emotion Generation*), and is expressed as follows:

$$f_{i,a}\left(m_j\right)=\begin{cases}\left(\dfrac{a_{i,upper}-a_{i,lower}}{m_{j,upper}-m_{j,lower}}\right)*\left(m_j-m_{j,lower}\right)+a_{i,upper}, & if\ ma_{ij}=1\\ 0, & if\ ma_{ij}=0\\ \left(\dfrac{a_{i,lower}-a_{i,upper}}{m_{j,upper}-m_{j,lower}}\right)*\left(m_j-m_{j,lower}\right)+a_{i,upper}, & if\ ma_{ij}=-1\end{cases} \quad (13)$$

where $f_{i,a}\left(m_j\right)$ is the linear mapping function relating mood *j* to activation point *a* for emotion *i*, $a_{i,upper}$ and $a_{i,lower}$ are upper and lower bounds for activation point for generating emotion *i*, $m_{j,upper}$ and $m_{j,lower}$ are upper and lower bounds for mood *j*, m_j is mood *j* current intensity, and $\overrightarrow{ma}$ is the mood-emotion activation point dependency matrix. This function produces the value for mood-based portion of the activation point variable, $a_{i,\ mood.}$

In **Figure 11**, the higher mood value results in a lower activation point, and consequently, in higher sensitivity to the emotion-eliciting stimuli, and vice versa. For example, if an individual in a highly negative mood encounters a disgust-eliciting stimulus, then a stimulus of a lower strength will be sufficient to activate the emotion of *Disgust*. Similarly, for an individual in a highly positive mood, a lower-strength positive stimulus will activate a positive emotion of *Joy*, compared to neutral or low positive mood.

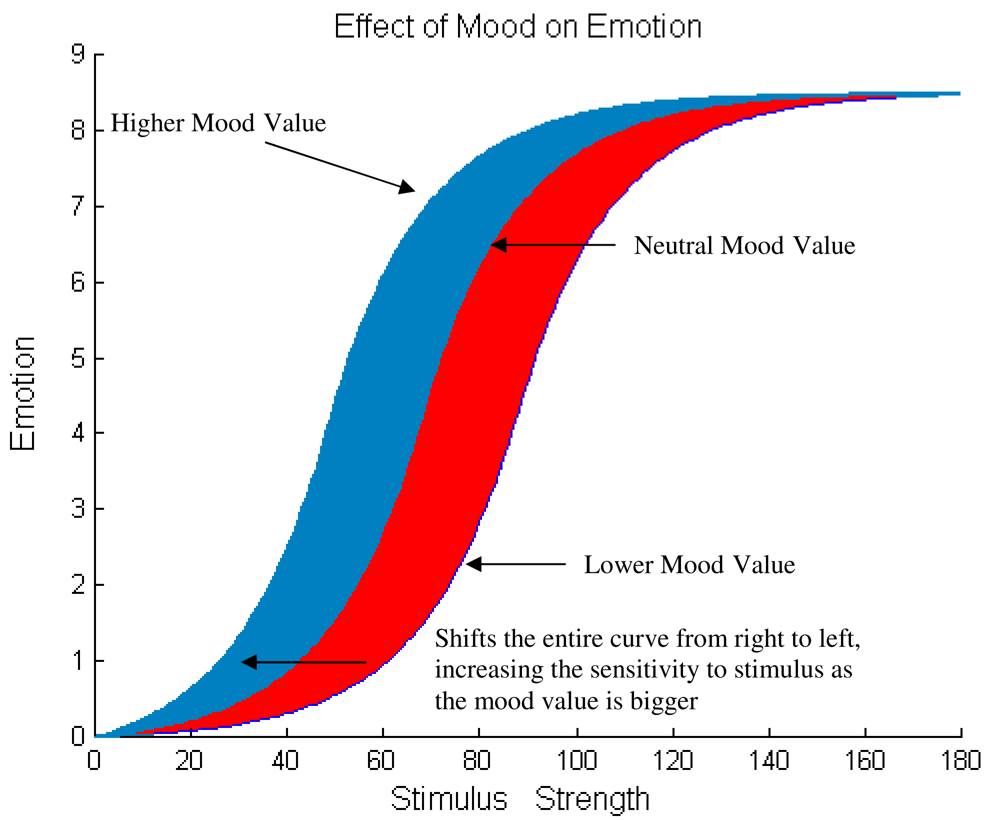


Figure 11: Effect of similarly valenced mood on the activation point of emotion generation. For example, higher negative mood will result in greater sensitivity to fear.

3.3.2.2.3 Combined Personality and Mood Influence

Finally, as both personality and moods influence the activation variable *a*, its overall value is the average of both personality and mood-based activation variables:

$$a_{i,\ overall} = (a_{i,\ trait} + a_{i,\ mood})/2 \qquad (14)$$

where $a_{i,\ overall}$ is the activation variable for generation of emotion *i*, $a_{i,\ trait}$ is the personality-based activation variable, and $a_{i,\ mood}$ is the mood-based activation variable.

Figure 12 shows graphically the combined effect of mood and personality on emotion generation. As previously, a higher value of a trait producing a direct influence on emotion results in a greater sensitivity to an eliciting stimulus, higher amplitude, and steeper slope; and a higher mood value for the same trait intensity results in yet greater sensitivity (lower activation point). For example, for a high-neuroticism individual in a predominantly negative mood, even a low-strength fear-eliciting stimulus will result in a significant emotional experience ("afraid of its own shadow" effect), whereas in a low-neuroticism individual in a positive mood a similar strength stimulus will hardly produce any fear.

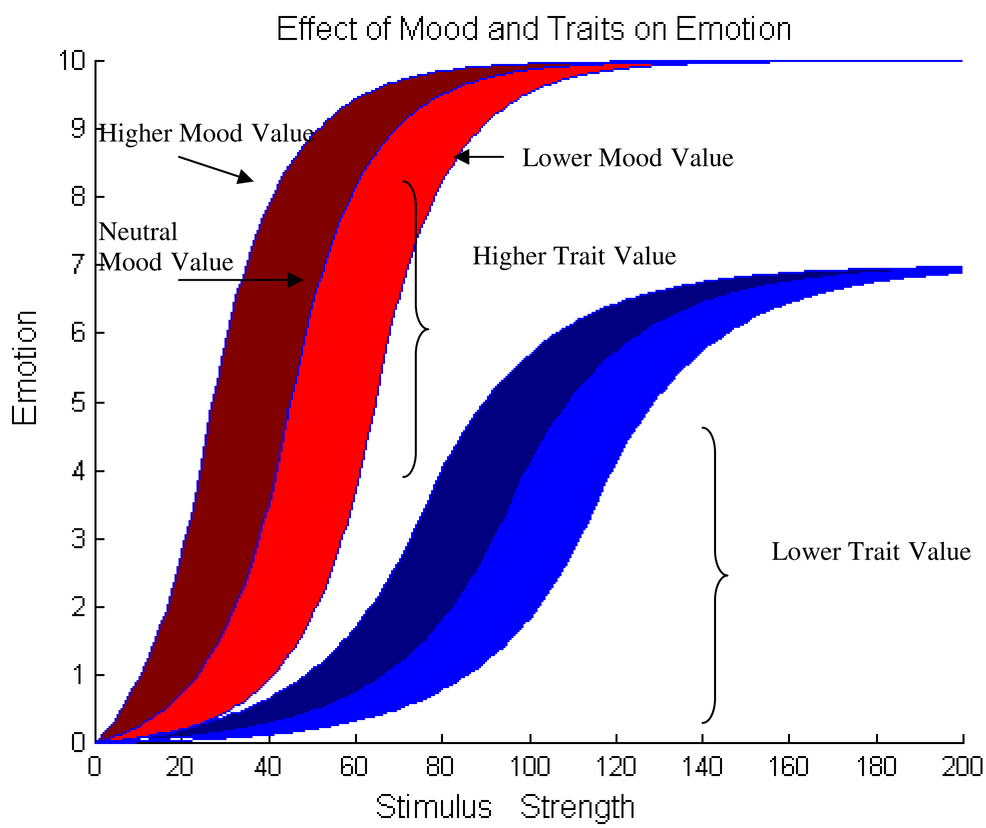


Figure 12: Combined Mood and Trait Influence on Base Emotion Generation

3.3.2.2.4 Attitude Influence on Emotion Generation

As was noted earlier in section 3.3.2.1, any attitude a robot might have towards an attitude-invoking object is taken into account while calculating the stimulus strength. Thus, instead of producing a change in specific emotion-generation variables (activation point, slope or amplitude), the intensity of an attitude will be reflected directly in generating a corresponding emotion.

3.3.2.3 Emotion Decay and Filtering

The base emotion generation described above took into account the current stimulus, and personality and mood influences. However, to account for short-term duration of emotions, their decay over time should also be modeled. Picard [20] compares an emotional response to ringing of a bell, where striking a bell initiates a response with a fast rise time and a more gradual decay. As a starting point, the following slowly decreasing exponential was chosen to model this gradual emotional response decay for this research:

$$E_{i,t,decay} = E_{t,base} - e^{(t-t_0)*d} \qquad (15)$$

where $E_{i,t,decay}$ is the intensity of emotion i at time t, t_o is the time at which emotion is activated (becomes greater than 0), and d is a variable that controls the rate of decay. Although it is not done at present, the decay rate variable could be different for every emotion, and could also be dependent on a particular personality trait, as response time (or decay rate) is likely to be influenced by intrinsic differences between individuals; for example, after a fear-provoking encounter, some people recover quickly, while others slowly [133].

Finally, in order to smooth the emotion change in cases of sudden appearance and disappearance of eliciting stimuli, a weighted averaging filter of the following form will be used:

$$E_{i,t,filtered} = (w_{current} * E_{i,t,decay} + w_{prior} * E_{i,t-1,filtered})/(w_{current} + w_{prior}) \qquad (16)$$

where $E_{i,t,filtered}$ is the final intensity of emotion i at time t after filtering, $w_{current}$ and w_{prior} are weighting variables controlling the relative importance of current and previous emotional states. This filtering function will help to account for short-term lingering emotions even after the eliciting stimulus has disappeared.

Figure 13 shows the effect of decay and filtering on emotion generation. For illustration purposes, as a stimulus of a medium strength (plotted in blue dashed line) suddenly appears, grows, continues at same strength, subsides, and finally just as suddenly disappears, base emotion grows according to the aforementioned base emotion generation function (plotted in green dotted line). Due to the emotion decay, the emotion begins to slowly decrease (black solid line) as the stimulus strength stays at the same level (suggesting a habituation effect). Finally, after the stimulus disappears, due to the filtering, the emotion still lingers for a short time (red dashed line).

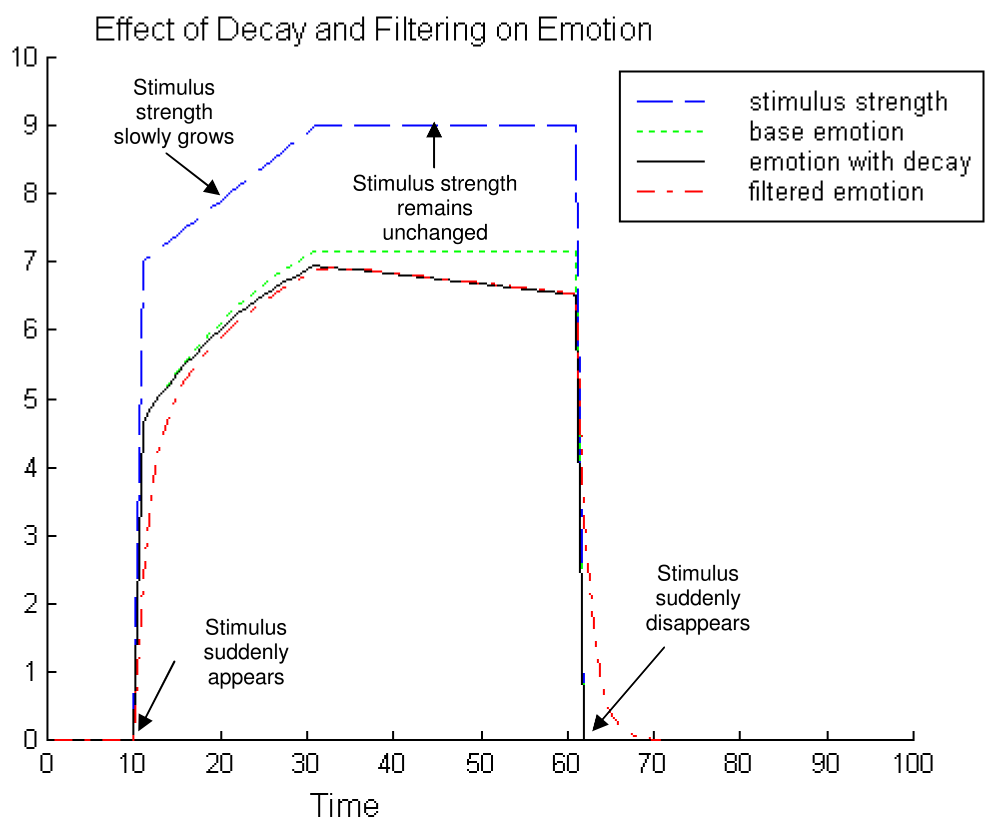


Figure 13: Effect of Decay and Filtering Functions on Base Emotion

3.3.3 INFLUENCE ON BEHAVIOR

Once the emotion is being generated dynamically with the presence of eliciting stimuli, how does it proceed then to influence the behavior? For this research, we plan to explore both the expressive/communicative function of emotion (especially significant for

HRI), and that of guiding action. By definition, emotion is a high activation state that may disrupt the current behavior and even hijack it if a strong threat to survival is encountered. Therefore, as an emotion is activated, it may in its turn drastically affect an active task behavior by changing the behavioral parameters. Such a behavior could a actually have been latent to begin with (i.e., with a gain of 0), but could be affected strongly enough that it will rise to dominate other behaviors feeding into the coordination mechanism. Although not addressed in this dissertation, an additional mechanism may also be developed by which an emotion of certain intensity can serve as a perceptual trigger for behavioral transitions (similarly to release mechanisms described in [135-137]).

3.3.3.1 Mapping from Emotions to Emotion-Specific Behaviors

As the proposed emotion-generation mechanism may be fairly complex, a simpler linear mapping from emotion to emotion-specific behavior parameters will suffice for the present, to be reexamined in future work. In order to obtain such a mapping, two sets of data points are required. The upper and lower bounds for emotions have been established earlier (0 is the lower bound, and the amplitude, g, is the upper bound). As with the task-specific behaviors, modified by traits, it is possible to establish both lower and upper desired values for emotion-specific behavioral parameters. For example, the highest desired value for a gain may be such that it could overwhelm all the other behavioral input in the coordination mechanism. By necessity, these upper and lower bounds should be determined empirically. Given these two sets of data points, we can fit them into a line of a general form of y=ax+b, thus providing a mapping from the emotion levels onto the emotion-specific behavior parameters. This provides a linear scaling mechanism for each parameter through linear functions that produce a line, where $f_{ij}(E_j)$ is the mapping function between emotion *j* and behavioral parameter *i*. Similar to traits, an emotion can have no, direct, or inverse influence on a behavioral parameter. A behavior-emotion dependency matrix $\overrightarrow{eb} = [eb_{ij}]$, where $eb_{ij} \in \{-1,0,1\}$ is defined; 0 refers to the absence of influence, +1 to direct influence, and –1 to inverse influence. The resulting emotion-based behavioral parameters will be used in the corresponding emotion-specific behaviors once they are activated.

3.3.3.1.1 Object Avoidance Gain Example

Consider an example of an object avoidance behavior. This behavior is similar to an obstacle avoidance behavior, but objects, as opposed to obstacles, are designated as emotion-eliciting stimuli (e.g., color-coded, with a color linked to a particular emotion). The magnitude of the object avoidance vector is defined as follows:

$$Obj_{magnitude} = \left\{ \begin{array}{ll} 0, & if \;\; d > S \\ \frac{S-d}{S-R} G, & if \;\; R < d \leq S \\ \propto, & if \;\; \mathrm{d} \leq \mathrm{R} \end{array} \right\} \qquad (17)$$

where *S* is default sphere of influence, *R* is radius of the object, *G* is the default gain, and *d* is the distance of the robot to the center of the object in question.

This particular behavior is a good candidate for fear response, therefore in this example we will map the emotion of *Fear* to Object Avoidance gain. *Fear* has a direct influence on the Object Avoidance behavior: the stronger the *Fear*, the stronger the avoidance response. Thus, as *Fear* grows, so does the Object Avoidance gain; similarly, as *Fear* subsides, so does the gain. In **Figure 14**, *Fear* intensity and Object Avoidance gain are plotted with the following upper and lower bounds: $lower_{fear}$=0; $upper_{fear}$=10; $lower_{Obj}$=0; $upper_{Obj}$=15; stimulus strength is shown for clarity.

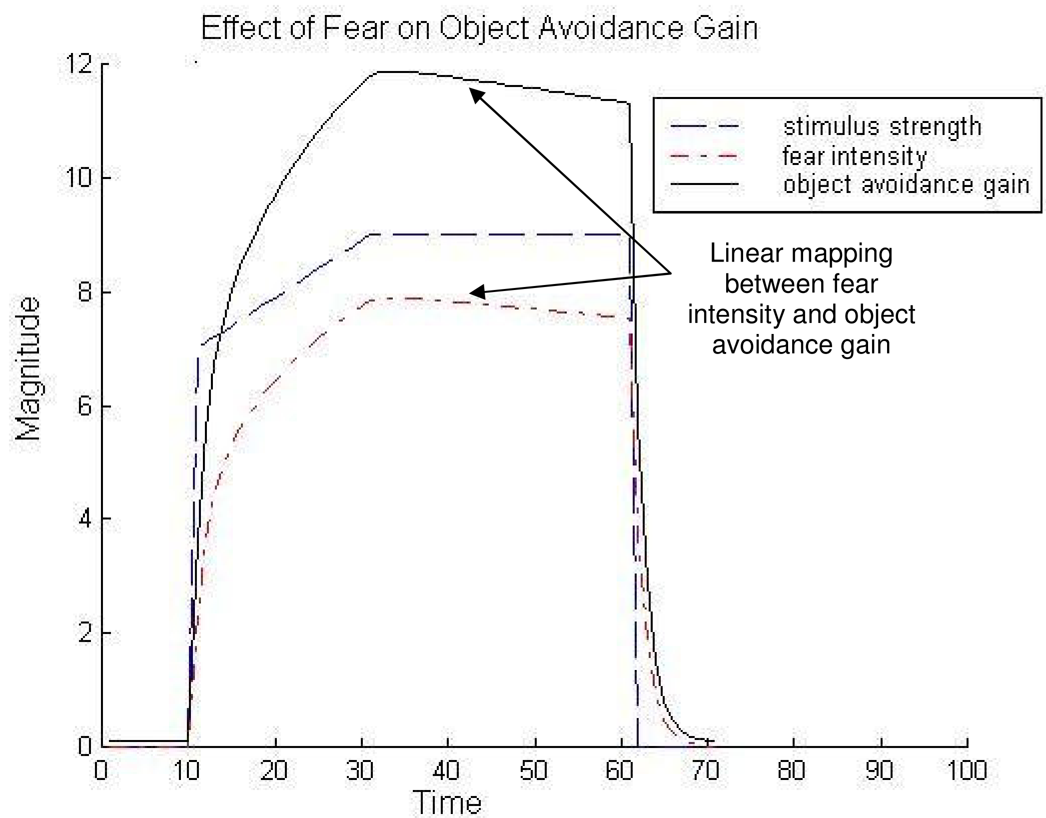


Figure 14: Example of Fear to Object Avoidance Gain Mapping

A similar emotion/behavior mapping mechanism can be used for specifying prototypical emotional expressions for humanoid robots as well. In particular, given the minimum and maximum allowable angles for certain joints, and minimum and maximum desired frequency for repeated gestures, we can provide a mapping between emotion strength and robotic gestures, movements, and body posture. For example, an expression of *Joy* often includes raising arms and shaking them; therefore, higher levels of *Joy* would result in a higher height to which the arms would be raised, and higher frequency of repeated arm movements, and vice versa for lower *Joy* levels. Finally, in cases where it is not feasible to provide a smooth, safe trajectory for body movements, a number of variations of gestures/body movements/posture can be coded a priori (e.g., those representative of Low, Medium, and High levels of *Joy*), and a simpler discretized mapping can be used. The latter approach is described in Chapter 5, *software* *architecture and implementation,* as applied to generation of emotional expressions in a small humanoid robot.

3.3.4 EMOTIONS SUMMARY

In this section, along with the psychological foundations which serve as an inspiration for the Emotion component, various aspects of emotion generation and its influence on behavior were discussed. In particular, mechanisms for stimulus strength calculation; for initial emotion generation based on stimulus strength, and personality and moods; and for emotional response decay were provided. The Emotion component is the most time-varying of all in the TAME framework. This variation is two-fold: 1) due to the time-varying nature of the eliciting stimuli; and 2) due to the response decay process. Stimuli can appear, stay present, and disappear at various points in time and will determine the change in emotion intensity. Emotions are by nature short-lived, therefore a separate time-based mechanism is used that allows emotions to subside even in the prolonged presence of stimuli. The overall result are short-term, high intensity spikes, usually rather separated in time, as only a small portion of objects serve as emotion-eliciting stimuli.

3.4 MOODS

Although references to moods permeate our daily lives, laypeople often use the terms “mood” and “emotion” interchangeably, to refer to a positively or negatively valenced state. However, there is a significant number of differences between these two phenomena, and in this section we will attempt to disambiguate moods from emotions, show their effects, and their potential usefulness in the robotics domain.

3.4.1 PSYCHOLOGICAL FOUNDATIONS

Both emotions and moods are generally viewed as affective states, where states refer to temporary conditions of an organism reflected in multiple systems [111]. The most recognized differences between these affective states lie along two dimensions, namely, duration and object specificity. In particular, moods last significantly longer than emotions, and provide an affective background, or “emotional color”, whereas emotions can be viewed as “phasic perturbations on this background activity” [10]. In accordance with this view is also that of Watson et al. [4], who views mood is a continuous variable affective state, or “stream of affect”. In terms of object specificity, emotions are object-focused, whereas moods can be often characterized as being diffuse or global [3, 111]. In addition, Watson et al. [4] suggest that mood represents a low activation state and is less intense, and thus expends less energy and bodily resources than emotion. Finally, as opposed to providing a fast, adaptive response to environmental contingencies, mood seems to provide “goodness” or “badness” information about a situation or environment [138], and consequently biases information processing [10].

3.4.1.1 Negative and Positive Affect

Watson et al. [31] suggest an approach that focuses on two broad mood factors: Positive Affect and Negative Affect, where Negative Affect refers to the extent to which an individual is presently upset or distressed, and Positive Affect generally refers to one’s current level of pleasure and enthusiasm. Isen [139] also contends that mild positive affect is not just an inverse of negative affect in its effects. Therefore, the moods will be represented along two different dimensions, positive and negative. The level of arousal for both categories can vary from low to high, which affects the nature of subjective feeling experience.

According to Watson’s theory, these two categories are not mutually exclusive and are largely independent of one another at moderate levels of affect; therefore, their combinations can form four basic types: high positive/low negative (e.g., happy); high positive/high negative (e.g., a mixture of fear and excitement, similar to feelings on a roller-coaster); low positive/high negative (e.g., depressed); low positive/low negative (e.g., disengaged). **Table 10** gives a comparative view of these factors and their level.

Table 10: Comparative View of Mood Levels

Valence	Level of Affect	Description
Positive	Low	Sluggish, Unenergetic, Disinterested in the Environment
Positive	High	Cheerful, Excited, Energetic, Alert, Self-Confident
Negative	Low	Calm, Relaxed
Negative	High	Nervous, Dissatisfied, Discouraged, Irritable

Although it may seem counterintuitive at first that a low positive mood value has a negative connotation (“sluggish”, “disinterested”), it refers to an insufficient level of pleasure and enthusiasm, rather than just a low level. Conversely, as negative affect signifies how upset or distressed an individual is, low level of negative mood would suggest that the individual is calm, rather than distressed.

3.4.1.2 Causes of Mood Changes

There are two main groups of factors that influence the current mood. The first one refers to an individual’s external and internal environment. Moods often follow a biological rhythm, and display variation dependent on time of the day, patterned variation across the days of the week, and seasonal fluctuations [4]. For example, energy levels, and consequently moods, may change throughout a day in a circadian cycle, which differs from individual to individual. An example of extreme seasonal fluctuations is Seasonal Affective Disorder, which may be linked to scarce sunshine in winter and fall [140]. From an evolutionary viewpoint, daily rhythmic mood changes may be partly due to the poor lighting conditions at night (which made any activity less successful and more dangerous). A body’s overall well-being is another factor that produces mood changes; lack of exercise, food or sleep and the general state of illness vs. health can influence the feeling of well-being. Finally, social interactions can also influence an individual's mood, and the amount of interaction needed would depend on the individual's personality [140].

Another group of factors refers to changes in short-term situational and environmental variables, including emotional episodes experienced throughout the day. Weather can be one of such variables; e.g., a dreary, rainy day can lead to a depressed mood, and a warm, sunny day may bring about a positive, happy mood [141]. A series of negatively valenced events can cumulatively produce a negative effect on mood, and vice versa, a number of pleasant interactions can shift the mood in the positive direction [10]. Many theorists also acknowledge the influence of emotions on moods, but the exact nature of this influence is unknown. Ekman [28] proposes that moods can be generated by "dense" emotional experience, where a specific emotion occurs with high intensity several times within a short time period. Another view is that every emotion tends to lead to general, diffuse responsiveness, and thus results in a mood change [3].

3.4.1.3 Mood Effects

Among various effects of moods on cognition, and, correspondingly, behavior, the following four have been repeatedly observed in a number of studies [142]. Not all of these effects are readily applicable to the robotics domain – they are presented here for completeness.

1. **Mood-congruent recall.** This finding suggests that while in positive mood, an individual finds positive aspects of a situation more salient, and therefore in subsequent recall these positive aspects feature more prominently. Similar effects are observed with negative moods, although to a smaller extent.

2. **Mood-congruent evaluation/judgment**. In accordance with the "feelings-as-information" model, people may judge a person, object or situation by asking

themselves “How do I feel about it?”, and may treat their general preexisting mood state as the answer to this question. Thus, more positive judgments would result under happy, rather than sad, moods [111]. This may be a significant determinant in attitude formation.

3. **Mood and systematic vs. flexible processing of information.** Schwarz [143] suggests that moods reflect the state of the environment: being in a bad mood signals a problematic situation, and being in a good mood suggests a benign one. It has been repeatedly observed that individuals in a negative affective state engage in a more systematic, analytical, and detail-oriented processing, perhaps in order to understand and correct for a failure or a problem. In contrast, good mood seems to engage more heuristic, routine, and general-knowledge processing, which may involve greater use of existing knowledge structures and stereotypes, and be more flexible [143]. In task-oriented situations, moods may serve as feedback about one’s performance, where positive affective cue may be perceived as an incentive, a “go” signal to pursue current inclinations, and negative affective cues as an inhibitive, or “stop” signal [111] . Thus success feedback should lead to the use of existing knowledge, whereas failure feedback should trigger learning and detailed analysis. Although this effect mainly refers to information processing, we apply it to behavioral styles, e.g., negative mood will cause the robot to behave more cautiously and systematically, whereas positive mood will encourage exploration and interaction. The reflection of the general situation through moods may also provide cues to humans about environmental/situational conditions, causing them, perhaps subconsciously, to reassess the situation.

4. **Mood and creativity**. According to Schwarz [143], creative thinking which involves exploration of novel situations and flexible use of diverse elements should be impeded by bottom-up, detail-oriented and narrow-focus style induced by negative moods, and facilitated by heuristic, flexible and top-down processing fostered by positive moods. Indeed, in various experiments, participants in a happy mood outperformed their sad counterparts in problem-solving and association tasks requiring creative solutions [144].

In addition, the Influence Infusion Model [145] suggests that moods have an effect on strategic social behaviors, namely, positive affect may bring about more confident, friendly and cooperative “approach” behaviors, whereas negative mood should lead to a more avoidant, defensive and unfriendly style.

For this research, we will explore, to a certain extent, mood influence on social behaviors, and also its effect on systematic vs. flexible behavioral styles.

3.4.1.4 Influence by Traits

Not only do people differ in their appraisal of situations, they also are different in their circadian rhythms, their ability to handle lack of sleep and food, the amount of social interaction they require, etc. It is not surprising, therefore, that each of the five personality dimensions is correlated with positive and/or negative affect. In particular, according to Watson et al. [4], Neuroticism, but not Extraversion, is strongly correlated with Negative Affect, Extraversion, but not Neuroticism with Positive Affect; Conscientiousness is moderately correlated with Positive Affect, and low negatively correlated with Negative Affect; and Agreeableness is negatively correlated with Negative Affect, and somewhat with Positive Affect. In addition, the Openness dimension is somewhat correlated with higher measures of both Negative and Positive Affect.

3.4.1.5 Influence by Emotions

As has been noted previously, emotional experiences bring about changes in the mood state, with positive emotions raising the scale of positive mood, and negative emotions causing the negative mood to rise. The resulting mood would depend on the current mood level, and the intensity of the same-valenced emotion experienced within a relatively short period of time (e.g., a day).

3.4.1.6 Implications for TAME

The aforementioned psychological findings allow us to make the following design decisions:

1. We will explore a number of types of mood change causes: internal well-being (e.g., battery level), environmental conditions (e.g., light level) and emotional influence.

2. The variable mood states are represented by two mood intensities: positive and negative, represented as two independent dimensions. The Mood component constantly monitors for any inputs the combination of which constitutes the base mood level.

3. The inputs to the module will consist of internal robotic states (e.g., battery level), environmental conditions (e.g., noisy sensor conditions), and emotion values.

4. The output of this module will consist of modified behavioral parameters (although an alternative approach was taken during the software design and implementation stage for computational simplicity, where moods produce a bias in trait values, influencing behavioral parameters in a more indirect way).

5. Emotions will influence mood change: positive emotions will affect positive mood, and the negative ones – negative. However, for this work we will leave out the

personality's influence on mood initially. The reason for this is that traits already influence a wide range of behaviors, including those affected by moods. As the primary influence of moods (as modeled here), is on the behavior of the robot, the effect of personality on moods will show indirectly through behavioral responses. In the future, we plan to identify other ways that traits could produce a difference in mood generation.

3.4.1.7 Application to HRI

Moods are by nature very subtle, but always present, and their effect on HRI is persistent, though less pronounced than that of personality and emotions. As a person's mood changes according to the situation or biological rhythms, changing the robot's mood accordingly may provide benefits not only in terms of decreased frustration and annoyance on the part of the person, but also in more objective task performance. For example, Nass et al. [146] explored the effect of affect in a car voice on driver's performance, and found that drivers who interacted with voices that matched their own affective state (happy and energetic voice for happy drivers, and sad and subdued voice for upset drivers) had less than half as many accidents on average. This application of moods to HRI is also in line with the similarity-attraction theory which predicts that users will be more comfortable with computer-based characters that exhibit properties similar to their own [147].

Robotic moods can also be used to induce a corresponding mood in humans, where such induction may help influence people's behavioral and attitudinal responses. For example, induction of Negative Affect may be desirable in dangerous situations, to direct a person's attention to the surroundings and prime them to be more receptive of a potential evacuation request by the robot (*co-protector role*). Induction of Positive Affect may be desirable in cases where people need to be encouraged to think creatively – e.g., in exploration tasks (*co-explorer and co-worker roles*).

Finally, expressive manifestation of robotic moods can alert a person to favorable or unfavorable changes in the environment or in the robot itself, especially if perception of these changes is based on sensor input not available through human senses. Consider the following scenario. A humanoid is guiding a human inspector through a partially secured search-and-rescue site, when the lights become dim. Although no immediate danger is visible, the robot's negative mood rises, and it displays the signs of anxiety and nervousness; no action per se is warranted yet, but the inspector, picking up the cues from the robot, becomes more alert and ready for action. This scenario is described in more detail in Chapter 6, where it was used for an HRI study.

3.4.2 REPRESENTATION AND GENERATION

Similarly to emotions, moods are not user-defined, but rather are updated continuously based on changing environmental conditions and a robot's internal state. The mood intensity matrix is defined as follows:

$$\vec{m} = [m_l],$$

where $-\infty < m_l < \infty$. Smaller mood intensities refer to "low positive" or "low negative" moods. As was mentioned earlier, low levels of negative mood are more desirable (they refer to being "calm, relaxed"), whereas low levels of positive mood are undesirable (they refer to being "sluggish, unenergetic, uninterested in the environment", lacking energy).

As there are only two dimensions of moods, positive and negative,

$$\vec{m} = \begin{bmatrix} positive(P) \\ negative(N) \end{bmatrix}$$

3.4.2.1 Mood Generation

Environmental, internal and emotional influences combine together to produce mood values. *Environmental influences* include level of light, noise level, temperature, humidity, and any other conditions that may affect proper functioning of a robot, depending on a particular environment (e.g., indoors or outdoors) and a physical platform. For HRI, *environmental influences* may also include any changes in the environment that may affect humans, even if they are not detectable by them (e.g., increased level of radiation or presence of potentially hazardous gas). *Internal influences* include battery (energy supply) level, internal temperature of the robot, sensor state (e.g., broken vs. in good working order). Some of these conditions – e.g., the light level outdoors or battery level – are cyclical, and can account partially for the cyclical nature of moods. For example, if light levels during day and night vary cyclically, so will the mood, as a function of light level. These influences together comprise *mood generation variables*. Finally, e*motional influences* represent the non-cyclical portion of moods, and are a part of a larger category of negatively or positively valenced events (the broader category may be explored further in future work).

In order to better define the role of these various influences on mood, it is helpful to define a neutral point for each environmental and internal variable. Such a point would represent absence of influence of the variable on the mood, e.g., if the battery is half-charged, then the mood is not affected. This neutral point may be different for different robots – e.g., the level of light sufficient to successfully recognize objects may be different depending on the types of cameras used. Depending on the direction of influence of these variables, below this point one of the moods will start to decrease, while the other will rise; similarly, above this point the opposite mood will start to rise, and its counterpart will decrease. Some of these conditions may affect only one mood, and not the other, in which case their influence is set to 0.

As no findings exist that would specify the nature of the relationships between moods and environmental and internal influence variables, as a starting point we will define the current level of mood as a weighted summation of these variables. Assuming that the same variables can affect both positive and negative moods, strengths of environmental and internal inf/uences can be represented in a matrix $\vec{l} = [l_i]$, where $0 < l_i < b_i$, where b_i is the hardware-dependent upper bound (e.g., battery level has a hard limit on how full it can be, etc.). The relative weights for each variable are stored in the mood generation matrix $\overrightarrow{mg} = [mg_{ij}]$. The values in this matrix are unit conversion factors, to convert the various mood generation variables to the same unit, and will be found experimentally for each variable. In addition, negative mg_i stands for inverse influence of the variable on the mood, positive mg_i stands for direct influence, and 0 signifies no influence. As moods are continuous, always present streams of affect, the base mood is continuously generated based on the current environmental and internal influences as follows:

$$m_{base} = \begin{bmatrix} m_{positive} \\ m_{negative} \end{bmatrix} = \sum_{i=1}^{N} \begin{bmatrix} mg_{i,positive} \\ mg_{i,negative} \end{bmatrix} \cdot (l_i - l_{i,neutral}) \qquad (18)$$

where $\overrightarrow{mg}$ is the mood generation matrix, $\vec{l}$ is the mood generation variable strength matrix, and N is number of mood generation variables. Although in theory, moods could range from negative infinity to positive infinity, in practice they will be theoretically bound by the (weighted) sum of the highest possible values of each mood generation variable.

Figure 15 illustrates an effect of one of the mood generation variables on mood. In this example, the cyclically varying light levels affect positive and negative mood intensities differently. It produces a strong direct influence on the positive mood, and so as the light level goes up, so does the positive mood level, with high amplitude. The light level produces a much smaller, inverse effect on the negative mood (e.g., it is more dangerous to move around in darkness).

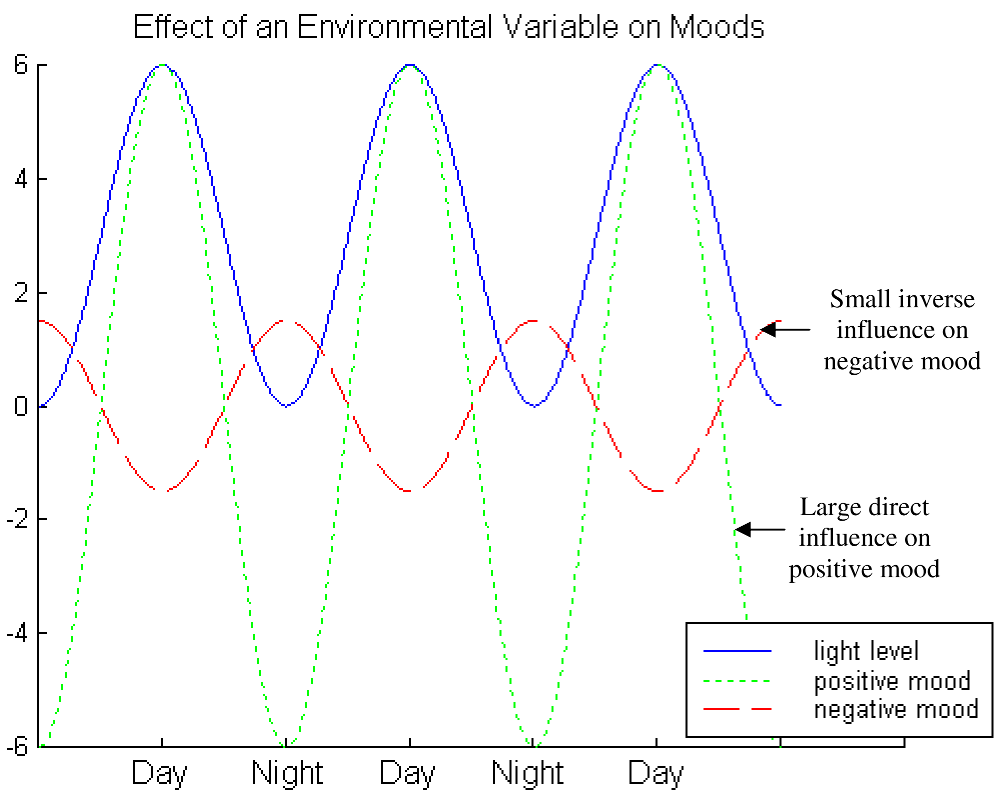


Figure 15: Different Effect of Light on Positive and Negative Mood Intensities

For the present, similarly valenced emotions will affect the corresponding mood intensities additively. Emotion intensities are given in the emotion intensities matrix $E = [E_j]$, defined in section 3.3.2, *Representation and Generation of Emotions*. The emotion influence adds to the existing base mood level in the following manner:

$$\begin{aligned} m_{positive} &= m_{positive,base} + E_k, \quad if\ E_k \in \{Interest, Joy\} \\ m_{negative} &= m_{negative,base} + E_k, \quad if\ E_k \in \{Fear, Anger, Disgust, Sadness\} \end{aligned} \tag{19}$$

where $m_{positive}$ is the emotion-based intensity of positive mood, $m_{positive}$ is the emotion-based intensity of positive mood, and $\vec{E}$ is the emotion intensity matrix.

Finally, as mood is a low-activation, slow-varying affective state, a filter will be used to smooth the influence of emotions and environmental and internal influences. In particular, using a filter to average the moods within the last 10-15 minutes will ensure that no particularly strong influences will be singled out and lead to overstimulating the moods.

Additionally, from an HRI perspective, in robots that are designed for sharing living conditions with humans for a prolonged time, circadian variations in mood may be introduced to provide mood congruency with the human, where user-defined cyclical daily, weekly and seasonal high and low points would be superimposed onto the base mood values. This method was explored in more detail by our research group in Park et al. [148].

3.4.3 INFLUENCE ON BEHAVIOR

Moods are mild by definition, and would only produce a mild effect, or a slight bias, on the currently active behaviors. Unlike emotions, which activate their own emotions-specific behaviors, moods will only influence active task behaviors. Unlike traits, which determine default behavior parameter values via scaling, moods produce an incremental effect on these parameters. Similar to traits and emotions, moods can have direct, inverse, or no, influence on a behavioral parameter. A behavior-mood dependency matrix $\overrightarrow{mb} = [mb_{ij}]$, where $mb_{ij} \in \{-1,0,1\}$ is defined, where –1 corresponds to inverse influence, +1 to direct influence, and 0 to absence of mood influence on behavior. Positive and negative mood may influence the same behavioral parameters, but most likely in opposite directions, thus to some extent canceling each other out. This combined influence is treated as additive for the present. As moods are updated continuously, new mood-based values of behavioral gains/parameters replace the existing trait-based values in the following manner:

$$B_{i,mood} = B_{i,trait} + K_i \sum_{j=1}^{N} mb_{ji} \cdot m_j \qquad (20)$$

where $B_{i,mood}$ is the updated behavioral parameter *i*, mb_{ij} is the mood-behavior dependency matrix value for mood *j*, m_j is the current value of mood *j*, N is the total number of mood categories (2), and *K* is a scaling factor. The scaling factor K is to be selected experimentally for each behavior parameter, to ensure that the moods produce only incremental effect as opposed to overpowering any of the parameters.

Figure 16 shows an example of incremental effects of moods on behavior. Suppose mood can bias the robot's obstacle avoidance behavior. For example, if visibility is poor, it may be advantageous to stay farther away from obstacles to accommodate sensor error, and vice versa, in good visibility it may be better to concentrate on task performance. Therefore, negative mood can bias the obstacle avoidance gain by raising it, and positive mood by lowering it. As established earlier, the trait of Neuroticism also affects obstacle avoidance by setting the default parameters to be used throughout the life-cycle, and the incremental effect of moods is shown against the space of trait-based defaults (plotted in solid blue line).

Alternatively, moods could affect behavior indirectly, through biasing personality traits. As personality traits specify behavioral parameters for the behaviors affected by moods, it may be computationally simpler to temporarily "change" the robot's personality which, in its turn, will bias the behavior via the mechanism described in section 3.2.2.2. This alternative solution has been explored by our research group in Park et al. [148], and was used during the software design and implementation stage (Chapter 5).

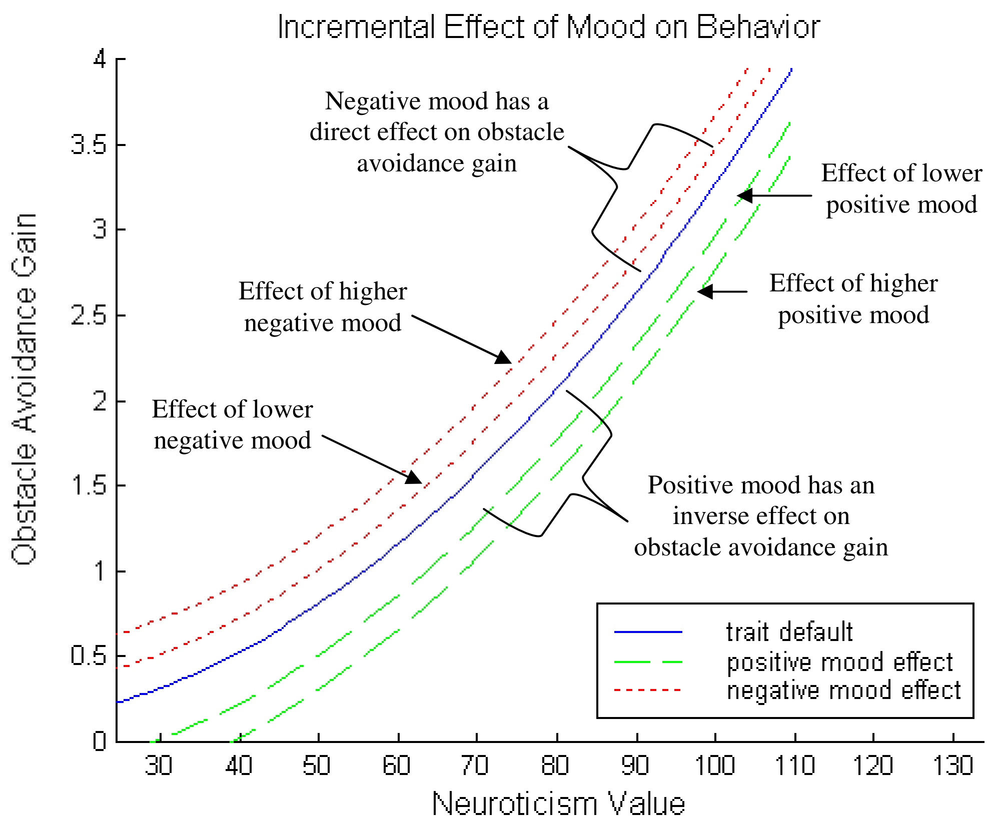


Figure 16: Direct and Inverse Effects of Moods on Obstacle Avoidance Gain at Different Neuroticism Values

3.4.4 MOOD SUMMARY

In this section, positive and negative dimensions of the Mood component from a psychological perspective, their update mechanism based on external, internal and emotional influences, and their effect on robotic behavior were discussed. In humans, moods vary cyclically with time of day, day of the week, and season. However, the underlying cause of such cyclical variation is not the passage of time per se, but rather environmental and internal conditions, such as changes in light intensity and energy expenditures. With this in mind, the mechanism for cyclical variation of moods over time in *TAME* will be dependent on such variation of the aforementioned variables. The slow, smooth nature of mood changes will then be achieved through a time-based filter in which previous mood states will be taken into account; the filter duration can be varied depending on the type of environment, physical platform and task requirements. The filter will also smooth out any drastic emotional influences.

3.5 AFFECTIVE ATTITUDES

Surprisingly, attitudes, such an integral component of human affective space, have been largely overlooked by the robotics community. People have an attitude towards everything – the latest political nominee, a distasteful flavor of ice-cream, their favorite pet, or a neighbor's annoying dog. If we intend for robots to be a part of our everyday lives, their sharing of our attitudes towards relevant objects around us would go a long way in establishing trust and communicational ease. For example, a child playing with a robot nanny or tutor may feel greater affinity towards the robot that acknowledges the child's likes and dislikes in toys and games. Our goal of providing affective attitudes for robots is two-fold. First, such attitudes would express a robot's current state in reference to a specific object or situation, and thus give a person cues of what the robot's impending behavior might be: for example, a robot expressing liking towards a certain object may be expected to approach and spend time with it. Second, robotic attitudes would reflect those of a human it interacts with on an everyday basis, thus building affinity and rapport in prolonged interactions. In this section, psychological foundations for the attitude component are described, and two alternative approaches for attitude generation, reflecting the aforementioned goals, are presented.

3.5.1 PSYCHOLOGICAL FOUNDATIONS

From a multitude of definitions of attitude, the following two provide the suitable emphasis for this research. Fishbein and Ajzen [8] define an attitude as a "learned predisposition to respond in a consistently favorable or unfavorable manner with respect to a given object". Another definition by Petty and Cacioppo [8] states that an attitude is "a general and enduring positive or negative feeling about some person, object or issue".

These definitions stress relative time-invariance (“enduring”), object/situation specificity, and the role of affect/affective evaluation in the attitude concept. In general, according to James Olson [149], most attitude theorists agree that: 1) evaluation constitutes a central part of attitudes; 2) attitudes are represented in memory; and 3) affective, cognitive and behavioral antecedents of attitudes can be distinguished, as well as corresponding consequences.

3.5.1.1 Representation and Structure.

Two potential ways of attitude representation in memory have been proposed: representation as knowledge structures, and as associative networks of interconnected evaluations and beliefs. The latter implies that elicitation of one attitude will make closely related attitudes more accessible.

Although no widely accepted taxonomy of attitude structure exists, the tripartite view of attitudes has been so far the most prominent [149, 150]. It suggests that attitudes can be formed based on affective information (accompanying affective states), cognitive information (knowledge-based evaluation), and behavioral information (influence of prior actions). Similarly, attitude-generated responses can be of affective (expression of like/dislike), cognitive (expression of beliefs), or behavioral (actions/action tendencies toward the target) nature. As this research is primarily concerned with affect, only the affective component of attitudes will be examined.

3.5.1.2 Functions of Attitudes.

A number of functions that attitudes may serve has been identified: adaptive or utilitarian function, knowledge or economy function, expressive or self-realizing function, ego-defensive and social-adjustive function [8, 151]. Adaptive function serves to guide behavior toward desirable goals and away from aversive events or actions. Knowledge function refers to managing and simplifying information processing tasks (e.g., stereotypes and prejudices). Expressive function refers to attitudes as a means of expressing personalities and values to others, and finally, ego-defensive function of attitudes serves to protect from self-threatening thoughts or urges (e.g., by denigrating others). Finally, social-adjustive function posits that attitudes facilitate the maintenance of relationships with others. Depending on the function a particular attitude serves, it may be formed and may influence actions differently. For our purposes, the most relevant functions are expressive and social-adjustive, as these are deemed the most useful for human-robot interaction: robot attitudes would both provide human participants with clues for impending robot behavior and help establish and maintain rapport. Both adaptive and knowledge functions would be advantageous to explore in the future: for behavior selection and regulation, and for facilitation of decision-making by reducing decision state space by automatically rejecting outcomes connected to undesirable entities/events (e.g., dislike or hatred), or providing incentive for choosing those connected to desirable entities/events.

3.5.1.3 Characteristics of Attitudes, and Attitude-Behavior Relation.

Attitudes can be characterized by their accessibility – the more easily accessible are the attitudes, the more influence they have; their strength (extremity) – similarly, the stronger the attitudes, the more direct influence on behavior they are capable of producing [149]; their complexity – the extent to which attitude-relevant information represents a number of distinct underlying dimensions; and their ambivalence - whether the attitude includes both negative and positive evaluations [151]. Accessibility refers to ease of retrieval of attitudes from memory, and the stronger the association between an attitude and its object is, the faster the attitude can be activated (associative networks representation may be particularly suitable to model this effect). The degree to which an attitude is held is represented by its strength, or extremity; strong attitudes resist change and produce widespread effects on both perception and behavior.

According to a number of theorists (Fishbein and Ajzen, Fazio [149]), attitude and behavior are correlated, and under certain conditions attitudes have a fairly high predictive power. Ajzen's [150] Theory of Planned Behavior suggests a causal chain from beliefs to attitudes to intentions to behavior. Attitude strength is determined as a function of the belief, i.e., the subjective probability that the object/stimulus has the attitude-inducing attribute (in practical terms, it could be translated into confidence level of smile recognition by a face recognition software). Attitudes in this model produce an indirect effect on behavior through intentions – action tendencies to behave in ways that are consistent with one's attitudes. The theory also proposes two other determinants of behavior: subjective norms (social approval of intended behavior) and perceived behavioral control. Although the direct effect of attitudes on behavior is beyond the scope of this work, the theory suggests directions for future research.

3.5.1.4 Attitude Formation and Antecedent Categorization.

Snyder and DeBono [152] identify three variables influencing attitude formation: individual differences/personality traits; target objects and their attributes (e.g., a smile); and situation types (e.g., stressful situations versus relaxed). As mentioned previously, cognitive (beliefs) and affective (affective states) processes, as well as past actions can serve as antecedents in attitude formation. It's been also shown that repeated exposure to a stimulus results in heightened positive evaluations [149], perhaps mediated by increase in positive mood due to such exposure.

According to Greenwald's Levels of Representation (LOR) system [153], there are five major representational levels: features (primitive sensory qualities), objects (a unified set of features), categories (groups of objects), propositions (abstract category types, such as action, instrument, target), and schemata (rule-governed groups of propositions). Attitude antecedents are argued to belong to the four out of five categories, excluding features; however, for this research, only objects and categories are easily applicable, and features are used as object components (attributes).

3.5.1.5 Influence by Emotions and Traits

Affective attitudes are closely related to emotions [3], and may even originate in a comparable emotion. From a computational perspective, if a certain object causes an emotion, then such emotion becomes one of the features of the object, and is used as an attribute that affects attitude formation. Negative emotions influence the attitude in the negative direction, while positive produce a positive influence.

The influence of personality traits on attitudes is multifaceted, but largely unexplored; therefore it will be left out of the framework for the time being.

3.5.1.6 Influence by Moods

It has been stated earlier (subsection *3.4.1.3, Mood Effects*) that current affective state influences evaluative judgments, and thus attitude formation; in particular, a person in a more positive mood will form a more favorable impression of a person or an object than one in a more negative mood. Clore [154] claims that the ease with which these affective biases can be shown suggests that people may often base their judgments and decisions on how they feel at the moment.

3.5.1.7 Implications for TAME

Taking into account the psychological foundations, we can make the following design decisions:

1. As we are only concerned with affective attitudes, and those have a strong connection to emotion elicitation, attitude component will have no direct influence on behavior. Rather, it will serve as a part of emotion generation, both in terms of helping to determine the type of emotion invoked, and its intensity. However, as

this work progresses, we plan to identify the areas of influence of attitudes on behavior and provide a mechanism for doing so.

2. We will explore three sources of attitude formation: attitude-inducing object attributes, current emotions, and the current mood state. Therefore, the inputs into the attitude component will consist of object attributes (features), robot's emotional state, and its mood level.
3. The attitude component will monitor sensor data for potentially attitude-inducing objects, and, if such are present, will output an attitude value for that object to serve as an input into the emotion component.

3.5.1.8 Application to HRI

The main application of affective attitudes to HRI lies in their potential to promote rapport, trust and attachment in prolonged interactions, in which robotic attitudes would reflect those of its interaction partner. For example, a child playing with a robot nanny or tutor may feel greater affinity towards the robot that acknowledges the child's likes and dislikes in toys and games.

3.5.2 REPRESENTATION AND FORMATION

Given the dual goal of providing robotic attitudes for HRI – expressing the robot's own attitudinal state in some cases, and the user's attitudes in others – we propose two distinct methods for attitude generation. The first one is more general, does not require input from the interacting human, and reflects a learned attitude a robot holds towards an object. The second method is based on the attitudes held by the interacting human, and therefore, the expression of resulting attitudes would be different for each user, even if the objects are the same. The former method is preferred for affinity building with a number of people the robot would be expected to interact with on an everyday basis.

3.5.2.1 Attitudes Reflecting the Robot's State

For an autonomous robot regularly operating among people it would be advantageous to express certain attitudes to surrounding objects, to both make it appear more life-like, and provide cues for behavior prediction. Such an affective attitude can be represented as a single value A, ranging from $-\infty$ to ∞, where 0 represents a neutral (or absence of) attitude, negative values represent an increasingly strong negatively valenced attitude (ranging from a mild dislike to hatred), and positive values refer to an increasingly strong positively valenced attitude (e.g., from a subtle like to adoration). Attitudes are object-specific, and an initial attitude for a particular object (*y*) would consist of a combination of positive and/or negative attributes (features) of this object. These object attributes or features are similar to those used in the calculation of stimulus strengths for emotion generation (subsection 3.3.2.1, *Eliciting Environmental Stimuli*), but reflect the more permanent properties, such as size, color, category, etc., rather than distance or orientation. Object attributes for attitude formation can be represented as a matrix $\vec{o}_y = [o_{iy}]$, where $-\infty < o_{iy} < \infty$. Such attributes are not limited to properties of the object only; for example, an emotion invoked by the object and any actions taken by the object may be considered "attributes".

The initial value of the attitude for object *y* ($A_{y,init}$) is calculated as follows:

$$A_{y,init} = \sum_{i=1}^{N} o_{iy} \qquad (21)$$

where $A_{y,init}$ is the newly-formed attitude for object *y*, o_{iy} is an attribute *i* of object *y* that is involved in the attitude formation, and *N* is the number of attributes for object *y*.

For purposes of illustration, suppose a robot encounters a small purple coffee table for the first time. The robot "loves" purple (color), "likes" furniture (category) and "dislikes" anything small (size); color and category for this particular object are positive attributes, and size is negative. Provided the attribute values are normalized to fit between -10 and 10, and the color feature is assigned 9, category 5, and size -5, then the attitude of the robot toward the newly encountered coffee table will be positive overall:

$$A_{coffeTable,\ init} = 9 + 5 + (-5) = 9$$

Assuming that initial impression is the strongest, substantial changes in attitude are fairly hard to achieve, therefore any subsequent exposure to the same object would result only in incremental change. This will be achieved through discounting any additional positive or negative object attributes (features) to a certain extent. The updated attitude value for object y for n-th encounter ($A_{y,n}$) would then be calculated as follows:

$$A_{y,n} = A_{y,n-1} + \lambda_n \cdot (\sum_{i=1}^{N} o_{iy}) \qquad (22)$$

where $A_{y,n-1}$ is the attitude towards object y at encounter n-1, n the total number of encounters up to date, $\vec{o}_y$ is the matrix of attributes for object y, and λ is the discount factor.

Following the same coffee table example, suppose the next time the robot encounters it, the coffee table has been painted orange instead of purple. If orange is a color the robot “dislikes”, (the color attribute = -6, size = -5, category = 5, and λ = 0.1), then the overall attitude will shift towards negative, but only slightly:

$$A_{coffeTable,\ 2} = A_{coffeTable,\ 1} + \lambda_2 * ((-6) + (-5) + 5)) = 9 + 0.1 * (-1) = 8.9$$

In the framework, the current mood states modify either an existing or forming attitude. According the finding on mood-congruent judgment, a positive mood increases the value of the attitude (a_y) towards an object y, and the negative mood – decreases it as follows:

$$A_{y,mood} = A_y + K(m_{positive} - m_{negative}) \quad (23)$$

where $A_{y,mood}$ is mood-enhanced value of robot's attitude towards object y, A_y is the original value of robot's attitude towards object y, $m_{positive}$ is the current positive mood value, $m_{negative}$ is the current negative mood value, and K is a scaling factor to bring moods and attitudes to the same units.

The mood effect is usually stronger when there are only few other attributes to influence the attitude towards an object. In some situations, this effect does not last beyond the situation itself, therefore it is important not to overestimate it by making it incremental.

Once the overall attitude for an object is calculated, it is forwarded to the Emotion component, where it serves as an attitudinal cue for stimulus strength calculation (subsection 3.3.2.1, *Eliciting Environmental Stimuli*); a negatively-valenced attitude would contribute to generation of negative emotions (fear, anger, sadness and disgust), and a positively-valenced one to positive emotions (joy and interest). **Figure 17** depicts the overall algorithm for computing attitudes reflecting the robot's state graphically. Given that this method of attitude generation is less relevant for human-robot interaction than the one reflecting users' attitudes (case-based method), and it presupposes object permanence (which is a hard problem in its own right), it was not explored further in this dissertation.

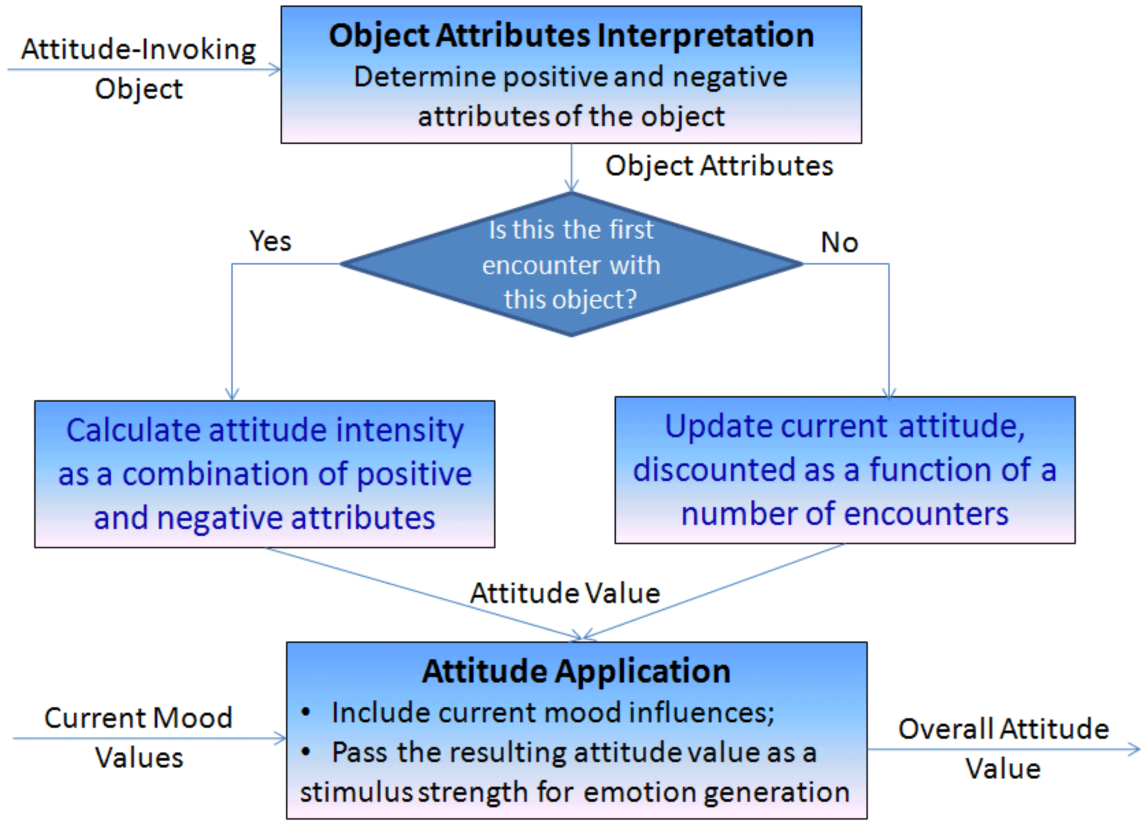


Figure 17: Process View for Formation of Attitudes Reflecting Robot's State

3.5.2.2 Case-Based Attitudes

For a robot involved in everyday interactions with a few of the same people in the same environment, it would be advantageous to learn those people's attitudes towards certain objects or even situations that may be repeatedly encountered, in order to promote bonding and rapport. With this in mind, the following properties of attitudes are taken into consideration: attitudes are learnable and are derived from experience (be it human or robotic), are widely variable across people, and they may be expressed as emotions towards specific objects, albeit more enduring and persistent. These properties make case-based reasoning (CBR) methods applicable for attitude formation, where an initial set of cases would be provided by each user, with a possibility of correcting erroneous attitudes for newly-encountered objects, and where the resulting attitudes would be expressed through corresponding emotions by the robot.

Instance-based reasoning, a highly syntactic CBR approach, is the most appropriate one for our work, as our aim is not problem-solving, and general knowledge would be of rather limited use, given the wide differences in attitudes between people. A flat index structure is employed, with apparent syntactic similarity only (without relying on general domain knowledge), relying on user interaction, especially in the beginning. Incremental supervised learning is used to adjust feature weighting scheme based on user input to learn relative importance of object features.

The case-based method for attitude generation follows the four steps typical of a case-based system [155]: 1) retrieve the most similar case or cases; 2) reuse the information in the case to solve the problem (display the attitude-based emotion from the best case); revise the proposed solution (based on user input); and 4) retain the parts of this experience likely to be useful in the future. The high-level process of using CBR for determining robotic attitudes is presented in **Figure 18**. The input data from the robot are interpreted into a feature vector for each object, which serves as an index for case retrieval along with a user ID; once an index is assembled, a similarity score is calculated for each case in the case base, and the best case is selected based on this metric; the case is then applied by providing the emotion component with attitude-based stimulus strength for each affected emotion; finally, if the user input is solicited, a new case based is added to the case library and feature weights are adjusted if the user disagrees with the displayed attitude.

As the total number of cases continuously maintained is not expected to be high (typically involving only a handful of people a robot would interact with in the same surroundings), memory, storage and computational complexity issues are not explored in this context.

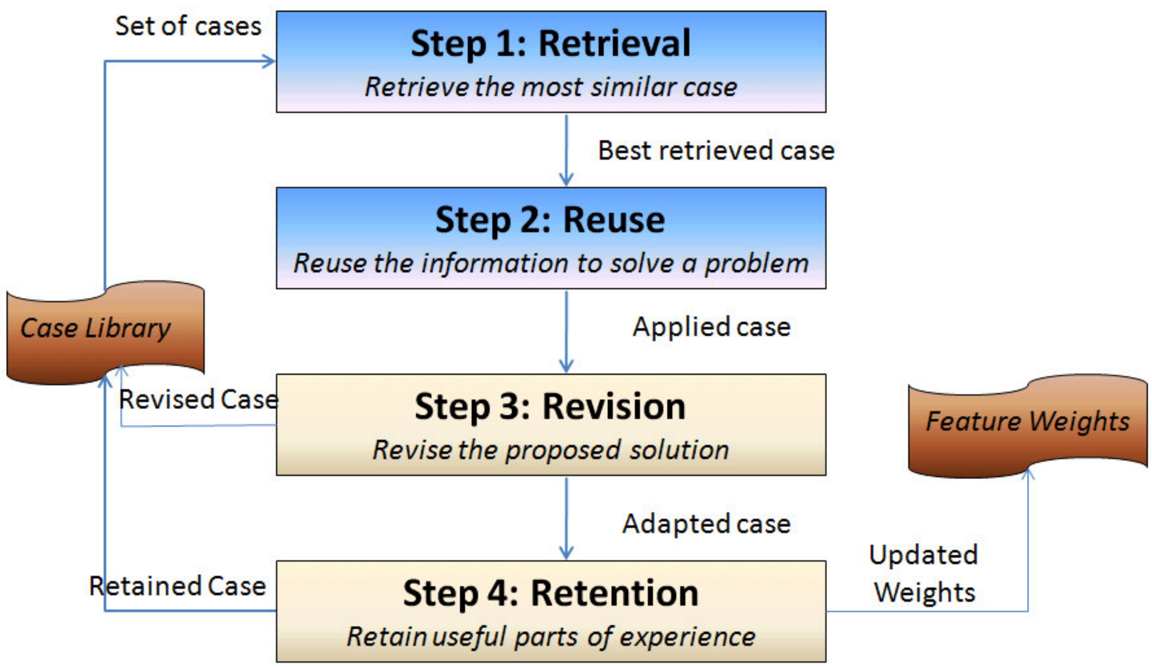


Figure 18: A High-Level Process View of CBR for Robotic Attitudes

3.5.2.2.1 Case Structure

A case library maintains a collection of cases containing information about user-specified or previously encountered objects along with attitudes towards them. A case *C* consists of an index vector *I*, and an output valence vector A_O of attitude-based emotion-stimulus strengths for an object *o*. The index vector is composed of User ID, to differentiate between users, a time stamp (for possible future linking attitudes and moods) and an Object Feature vector *OF* (20):

$$\begin{aligned} C &= [I, A_O] \\ I &= [userID, timestamp, OF] \\ OF &= [size, color, shape, material, category, etc.] \end{aligned} \quad (24)$$

Inputs from the robot include user ID, and a set of object features, either derived directly from raw sensor data, or interpreted from object identifiers, as is more often the case with current robotic systems. Object features refer to certain attributes an object might posses which would be relevant for attitude formation, for example, *color, size, material, shape and category/object type*. These features can be either continuous numeric values or categorical features, for example, wood, metal, glass or cloth are possible values for the *material* feature, and cubes, spheres and cones are possible values for the *shape* feature. Sometimes, certain features may not be applicable to an object (e.g., cars are not generally described in terms of cubes or cones), and sometimes, certain features may be hard to discern (e.g., it may not be possible to figure out whether a moving car has a cloth or leather interior, or a color may not be discernable in the dark); in both of these cases a feature is characterized as 'Not Applicable/Not Apparent'. In other cases, certain features may not play a role in attitude formation for a certain user; in this case, they are characterized as 'Don't Care'. These special values are used differently during similarity score calculation. Once object features are identified, they are combined with a User ID into an index for case retrieval.

The output vector represents an attitude towards object *o* and is defined as:

$$A_O = \{Fear, Disgust, Anger, Sadness, Joy, Interest\}, A_O \in \{0 - S\} \qquad (25)$$

where S is the maximum stimulus strength.

The output for attitudes represents attitude-based emotion stimulus strengths. This allows for an emotion to be generated based on an experienced attitude alone. The positive and negative aspects of attitudes (e.g., love or hate) are expressed through corresponding positive (joy or interest) or negative (sadness, fear, disgust, and anger) generated emotions. This representation allows to model attitudinal ambivalence – mixed feelings some people might have towards certain objects or issues, which contain both negative and positive evaluations.

Sample cases and object feature descriptions can be found subsequently in subsection 5.2.2.1.1., *The Attitude Component Implementation and Testing*.

3.5.2.2.2 Retrieval

During retrieval, first each case is assigned a score which reflects how similar this case is to the new object encountered by the robot. Based on this similarity score, a set of best-matched cases is retrieved (initial match). Finally, the best case is selected using a number of methods. The retrieval process is depicted graphically in **Figure 19**.

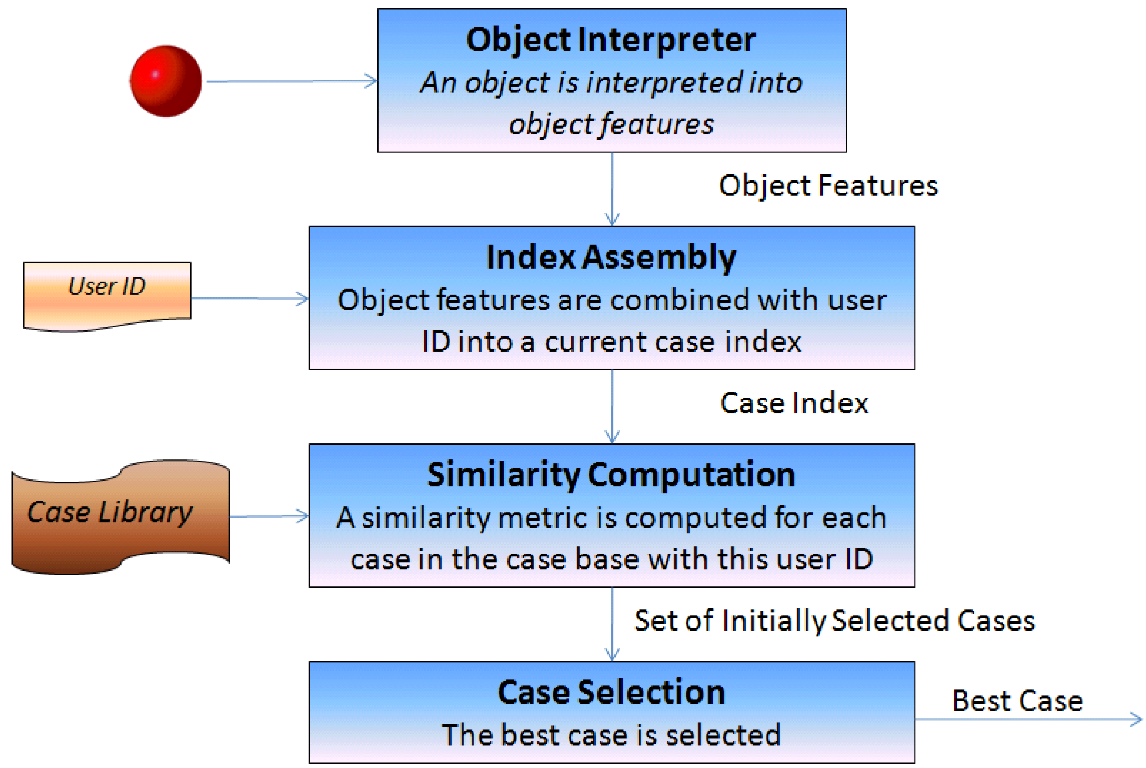


Figure 19: Retrieval Step

1. Similarity Metric

For this work, we are using syntactic similarity assessment to provide a global similarity metric based on surface match; this method has an advantage where domain knowledge doesn't exist or is difficult to acquire [155]. The similarity score is assigned to each case to determine a set of cases for the initial match; only the cases with a similarity score above a certain threshold are included into the initial match set. It is from this set that the best matching case is chosen. The threshold is determined experimentally, and may be different during the weights learning stage to draw from a wider base.

To obtain a similarity score for each case for a particular user, a Manhattan distance calculation is used to compute the shortest distance between the two vectors of Object Features (for the current object, and for each case in the library). The case with the shortest distance receives the highest similarity score. As not all features may be equally important in determination of attitude, each feature is weighted to indicate its relative importance; these weights are the same initially, but are learned incrementally from user input throughout the interactions. The similarity score between a new case and each existing case in the library is calculated as follows:

$$S(C_{new}, C_{existing}) = \frac{\sum_{i \in P} s_i * w_i}{\sum_{i \in P} w_i} \quad (26)$$

where s_i is a similarity score for feature i in the set P of predictor features (features valued 'Don't Care' or N/A' (not applicable) are excluded from this set); and w_i is a weight for the feature i.

A similarity score for each feature is calculated as follows:

$$s_i = \begin{cases} 1 - \left(\frac{|V_{i,new} - V_{i,exist}|}{R_i}\right), & if\ V_i is\ numeric \\ 1, & if\ V_{i,new} = V_{i,exist} \text{ and } V_i \text{ is a discrete category} \\ 0, & if\ otherwise \end{cases} \quad (27)$$

where $V_{i,new}$ and $V_{i,exist}$ are values of feature i for new and existing cases, respectively, and R_i is the range for numerically valued features for normalization purposes.

As was mentioned earlier, features valued as 'Don't care' and 'N/A' are excluded from the set P of predictor features used for calculation of the similarity score. However, excluding 'Don't Care' feature from the calcultion would create an overgeneralization bias, as the remaining features would be naturally weighed more heavily than otherwise, thus favoring cases with fewer features present. In order to reduce this bias, we use a discount factor γ for each feature valued 'Don't Care'. Thus, for each 'Don't Care', the overall similarity score for an existing case would be reduced as follows:

$$S(C_{new}, C_{existing})_{discounted} = S(C_{new}, C_{existing}) - S(C_{new}, C_{existing}) * \gamma * N \quad (28)$$

where $S(C_{new}, C_{existing})_{discounted}$ is the new discounted similarity score; $S(C_{new}, C_{existing})$ is the initial, non-discounted similarity score, γ is the discount factor (as γ gets smaller, more general cases will receive a higher similarity score); and N is the number of feature pairs valued 'Don't Care'.

Additionally, each case with a non-discounted similarity score above a predefined threshold is assigned a feature ranking, where the cases with the least number of 'Don't Cares' get the highest ranking (e.g., a case with no 'Don't Cares' will be ranked as (1). This is an alternative method for controlling the generality/specificity issue.

2. Final Case Selection

Once the initial matching is performed and the set of best cases is identified (those with a similarity score above an experimentally defined threshold), the best case is selected. A number of methods for selecting the final case are provided, of which the most appropriate to a particular situation should be used:

- The case with the highest similarity score is selected. This works well for a set where a clear winning case is present (the top score is much higher than the rest).
- A randomized roulette algorithm which probabilistically favors higher scores. This algorithm addresses the local minima problem, and is especially useful for cases with close similarity scores.

- Selection can be based on feature ranking, where a case with the highest score within the highest ranked cases is selected. This method is an alternative to the discounting method used for calculating similarity score for feature vectors with 'Don't Cares', and should be used in its stead.

A graphical view of the final case selection process is provided in **Figure 20**.

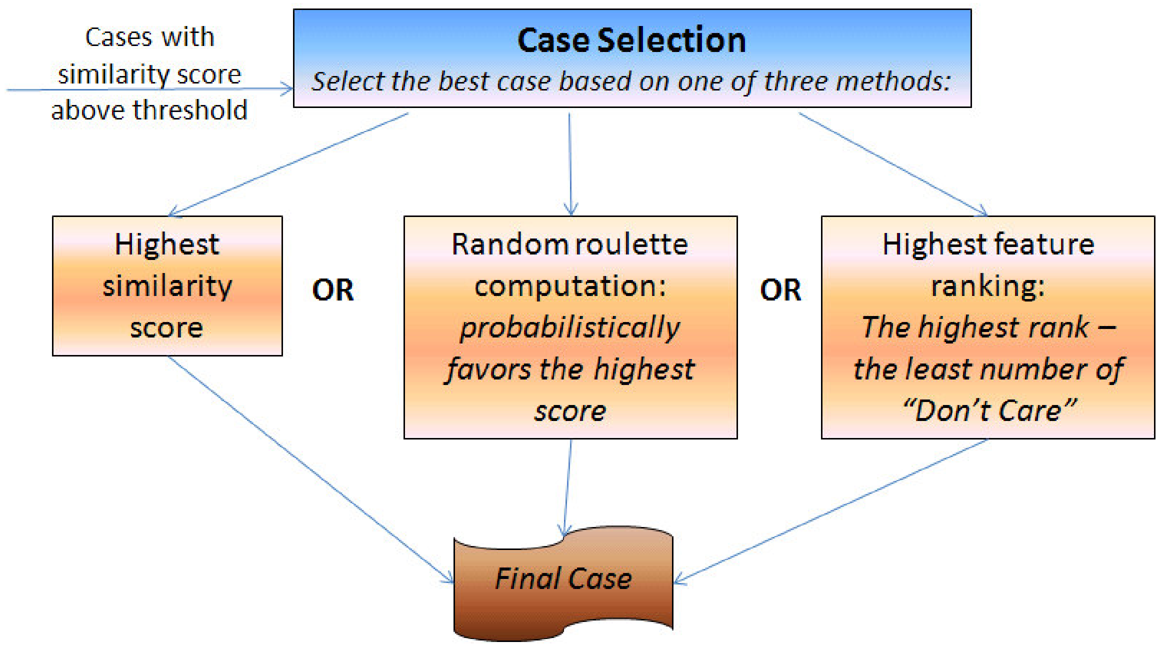


Figure 20: Final Case Selection: Alternative Methods

3.5.2.2.3 Reuse

As the attitudes in TAME result in a corresponding emotion expressed by the robot, once the best case is selected, its output is passed on to the Emotion component. The highest (dominant) attitude-based stimulus strength is then used to generate a corresponding emotion; as stated earlier, the output of the Attitude component both specifies the emotion to be generated, and the stimulus strength for it. If this particular object is also emotion-inducing, irrespective of any attitudes, then the output constitutes the attitudinal cue for overall stimulus calculation for generating corresponding emotions (subsection 3.3.2.1, *Eliciting Environmental Stimuli*).

3.5.2.2.4 Revision

Naturally, however, the best case selected and applied by the robot may not reflect a user's current attitude. In order to identify the attitude a user actually has for the presented object, a simple graphical user interface (GUI) can be utilized to solicit such input (see subsection 5.2.2.1.1, *The Attitude Component Implementation and Testing,* for examples). If revision/retention phases are desired (during supervisory learning stage, or when a new object is introduced), a user is asked whether the emotion displayed by the robot reflects what he/she may feel for that object. If the answer is "Yes", then the selected case is deemed adequate, and a new case is formed which combines the Object Feature vector from the new object, and the Output vector from the applied case (unless a case with exactly the same object features already exists, in which case no changes are made). If the user disagrees with the robot's emotional response, he/she is then asked whether the presented object is liked or disliked. After like/dislike selection, the user is asked to mark the intensity of each invoked emotion (Fear, Disgust, Anger and Sadness for dislike, and Joy and Interest for like) on a likert-type scale. To reduce the workload for the user, the information regarding the relative importance of features for determining the output is not requested; rather, the feature weights are incrementally learned during the retention phase.

There could be a number of reasons for a discrepancy between the final selected case and the user's input:

1. The new object may be too dissimilar to any of the cases in the case library.
2. The similarity metric was not sensitive enough to pick out the best matching case.
3. Even if the object has exactly the same features as a prior case, the user may have changed his/her attitude since the case library was initially filled.

Reasons 1) and 2) are addressed during the retention phase. Reason 3) suggests that a revision of a currently existing case is required. Assuming that an initial impression is the strongest, substantial changes in attitude would be fairly hard to achieve, especially for a strong original attitude [151], therefore any subsequent exposure to the same object would result only in incremental change. Therefore, the output of an existing case for which the user suggested a different attitude is modified only slightly, by adding to or subtracting from the original value a fraction of the value provided by the user:

$$A_{i,rev} = \begin{cases} A_{i,exist} + A_{i,user} * \gamma, & if\ A_{i,exist} < A_{i,user} \\ A_{i,exist} - A_{i,user} * \gamma, & if\ A_{i,exist} < A_{i,user} \\ A_{i,exist}, & if\ A_{i,exist} = A_{i,user} \end{cases} \tag{29}$$

where $A_{i,rev}$ is the revised value of attitude output i, $A_{i,exist}$is the existing value, $A_{i,user}$is the new user-supplied value, and γ is the discount factor, by which the influence of the new input is reduced. The resulting values are kept within the output bounds.

Once the new values are calculated for the entire Output Vector, the existing case is updated in the case library. **Figure 21** presents the flow of the Revision step graphically.

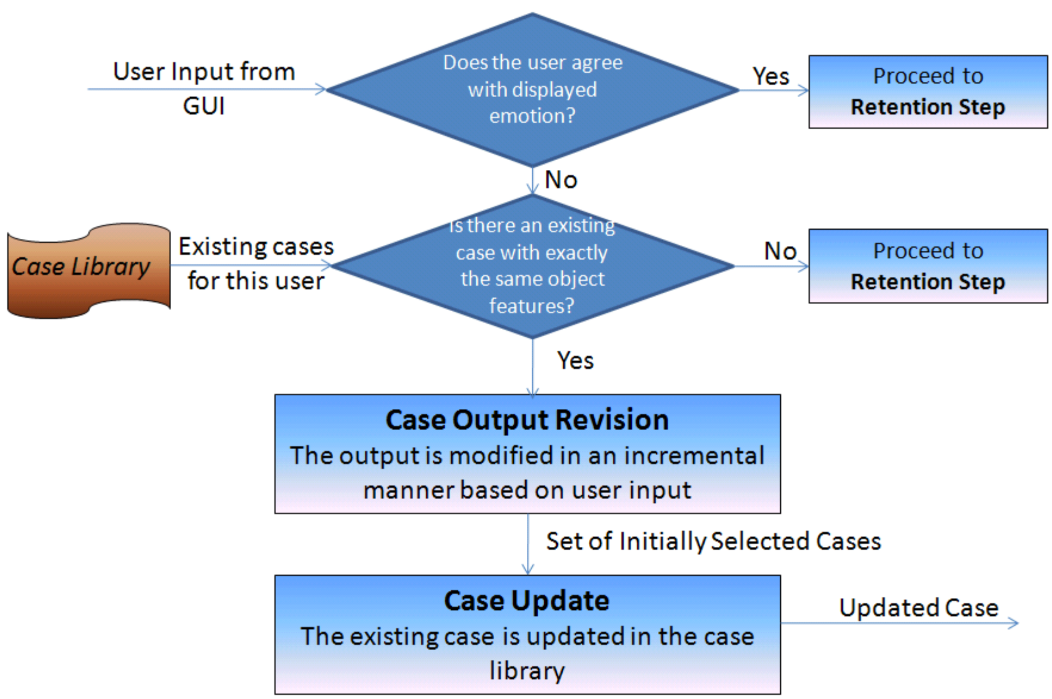


Figure 21: Revision Step

3.5.2.2.5 Retention

After user input is received, unless a case with the same feature values already exists, a new case is formed and added to the case library under a new case number and the user ID identifying the interacting user. The new case contains the Object Feature Vector for the newly presented object, and the Output Vector - provided by the user in case of disagreement with the robot's expression, or copied from the final selected case if the user agrees with it.

One way to improve the similarity metric and thus the overall performance of the system is to incrementally adjust feature weights to change the relative importance of each feature involved in similarity computation. These weights should be adjusted for each user, as people may have different preferences, and won't attach the same importance to the same features. Every time a user specifies a different output, an object feature needs to be identified that played the most important role in determining this output. In order to identify such a feature, the set of best cases selected originally is examined – the initial match, consisting of a number of cases with a similarity score above a certain threshold. First, from this set, only those cases are selected where output has the same dominant emotion as the output specified by the user. Then, a dissimilarity score is calculated for every feature across the set of cases with the same dominant emotional output. This calculation is similar to the one used during retrieval, except the scores are calculated per feature, and not per case, as we are interested in the features that contributed the most to produce the desired output. The average dissimilarity score D is calculated for each feature as follows:

$$D\left(F_{new}, F_{existing}\right) = \frac{\sum_{i \in C} d_i}{N} \qquad (30)$$

where d_i is a dissimilarity score for feature *i* in set C of cases with the same dominant output and N is the number of features in the set; features marked as 'N/A' or 'Don't Care' are excluded from the calculation.

Dissimilarity d_i for each feature *i* per case is calculated as normalized Manhattan distance:

$$d_i = \begin{cases} \frac{|V_{i,new} - V_{i,exist}|}{R_i}, & if\ V_i is\ numeric \\ 1, & if\ V_{i,new} = V_{i,exist} \\ & and\ V_i\ is\ a\ discrete\ category \\ 0, & if\ otherwise \end{cases} \quad (31)$$

where $V_{i,new}$ and $V_{i,exist}$ are values of feature i for new and existing cases, respectively, and R_i is the range for numerically valued features for normalization purposes.

Finally, the feature with the smallest dissimilarity score is deemed to be slightly more important than the other features for this particular user. In the case of a number of features with the same dissimilarity score, the feature shared by the largest number of cases is the best predictor of the output. The weight for this feature is then incremented, and the weights table is updated:

$$w_{best,updated} = w_{best,existing} + w_{best,existing} * \delta \quad (32)$$

where $w_{best,existing}$is the existing weight for the best case, and δ is the increment factor.

The next time the robot is presented with a new object, the learnt weighting scheme is used for calculating the similarity scores. **Figure 22** depicts the retention step graphically.

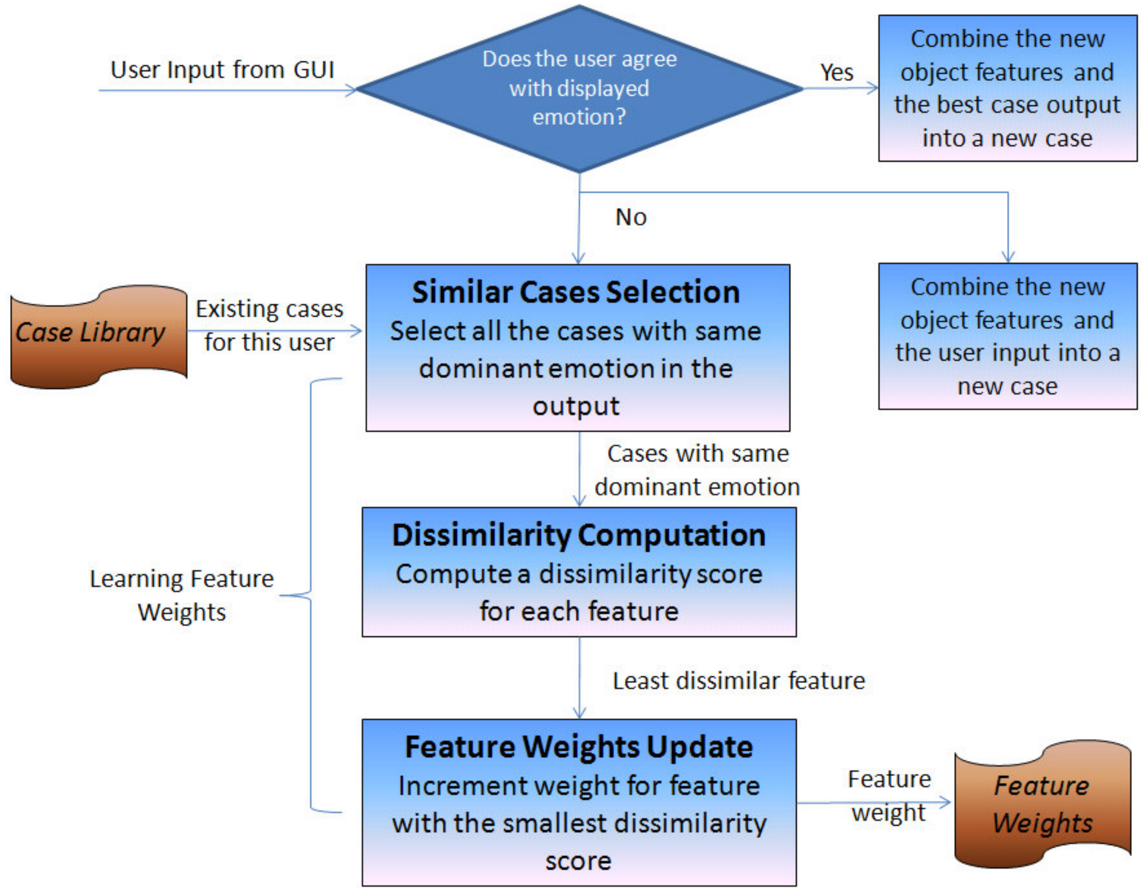


Figure 22: Retention Step: Retaining New Cases and Learning New Feature Weights

3.5.3 ATTITUDES SUMMARY

This section described the psychological basis behind the Attitude component, and two distinct mechanisms for attitude generation – one reflecting robot's attitudes irrespective of its social surroundings, and the other adjusting to its interaction partners. For the present, attitudes are modeled as persistent across time: once an initial attitude is formed, it stays the same unless the object of attitude is encountered again. In such a case, the attitude is reexamined, but the change is only incremental and is discounted as a function of the number of encounters. In the future, slow "fading" of attitudes over time will be examined, where the attitude intensity will decrease with prolonged absence of the attitude-invoking object ("out of sight, out of mind" phenomenon). The issue of "forgetting" old cases which are no longer appropriate for new environments has been previously explored in the robotics domain by Kira et al. [156], and the proposed solutions could be used as a starting point for designing a mechanism for attitudinal "fading".

3.6 CHAPTER 3 SUMMARY

In this chapter, the first subquestion of the main research question has been addressed: *"How can each of the aforementioned phenomena be represented, generated, and applied to robotic behavior, and what are the interactions between them that can provide additional benefit beyond that of each individual component?"* In particular, an overview of the *TAME* framework was presented, including psychological foundations for each of its components (sections 3.2.1, 3.3.1, 3.4.1 and 3.5.1, for each of the affective phenomena), their mathematical representation, mechanisms for their generation and influence on behavior (sections 3.2.2, 3.3.2, 3.4.2 and 3.5.2), and the interactions between them (throughout the chapter). Of these components, traits and attitudes constitute time-persistent dispositions towards behavioral patterns and emotionality (traits), and objects (attitudes). Traits do not change across the robot's life-span; here, life-span can be defined as referring to the duration of a single mission; a period of time during which the robot is expected to perform a certain type of task (this may span across multiple missions); or physical life of the robot, if its capabilities make a certain personality configuration particularly beneficial. Once specified, the trait combinations remain time-invariant until the human designer intervenes by changing the robot's "personality"; consequently, their influence on behavior and emotions will not change until such an intervention. Unlike traits, attitudes can change, albeit slowly. An initial attitude is formed when a corresponding object is encountered for the first time and remains the same until the next encounter. At that time, the update in either direction, towards either affection or dislike, is small. The change over time in attitudes will then depend on multiple encounters the objects for which an attitude already exists, rather than the passage of time per se. Alternatively, a robot can share the attitudes of those people it continuously interacts with, by maintaining a case base of attitudes for each interaction partner.

Unlike traits and attitudes, which are persistent over time, moods and emotions constitute the robot's ever-changing affective state. Moods can be described as continuous "streams of affect" [4], and, unlike emotions, can be neutral (neither high, nor low), but never altogether absent. Cyclical variations in moods over time will be determined based on the underlying variations in environmental and internal conditions; any sudden changes will be smoothed out by taking into consideration a number of prior mood states – filtering over a longer period of time will result in slower and smaller mood changes. Filtering will also help tone down any drastic spikes in the mood due to emotional episodes. Mood intensities will be continuously updated based on the current emotional state of the robot, in addition to the environmental and internal conditions; the most recent values will be used in calculating behavioral parameters and influence on emotion generation. Finally, as emotions are reactions to significant events (such as, for example, appearance of a dangerous object), they are expected to be relatively rare, but occur with high intensity and high impact. The changes in emotion intensity are the most immediate among those in all other components: emotions can go from nothing to very high in fractions of a second, provided a high-strength stimulus and suitable trait configuration; they can similarly drop just as fast with the stimulus's disappearance. A decay function is provided to account for habituation effects; a filter is used to help oscillations due to irregular perceptual input and to provide "lingering" emotional effect if the stimulus disappears very suddenly.

A summary of time-varying aspects of each *TAME* component is given in **Table 11**:

Table 11: Summary of time-varying aspects of the TAME components

	Traits	**Attitudes**	**Moods**	**Emotions**
Duration	Life-long	A few days to a few years	A few hours to a few weeks	A few seconds to a few minutes
Change in Time	Time-invariant	Persistent across time; change slowly with the number of times an object of attitude is encountered.	Change cyclically as a variable of time-varying underlying environmental and internal influences; any drastic changes are smoothed by taking into account previous mood states across a period of time; mood states vary faster than attitudes, but slower than emotions.	1) emotion intensity changes as eliciting stimuli appear, disappear, and get closer or farther away; occur as short-term peaks in response stimuli. 2) response decay describes decay of emotion as the time passes, even in the presence of stimuli.

The differences between the affective components in *TAME* are numerous and multi-faceted and each component provides a unique advantage. *Traits, Emotions, Moods and Attitudes* differ in their psychological, cognitive, and behavioral functions; their antecedents and generation mechanisms; their influence on behavior and application for human-robot interaction; and their duration and changes they undergo with time. Individually, each of them has a notable yet limited potential for robotics. Together, they provide a stepping stone for transforming machines into companions, and creating a richer and more vibrant illusion of life.

4 EXPLORATORY EXPERIMENTAL STUDY

In order to explore the issues of feasibility and potential usefulness of the *TAME* framework, as well as to inform the design of the software architecture incorporating the computational model, a longitudinal human-robot interaction (HRI) experimental study has been designed and conducted. Expressions of emotions of Joy, Interest, Anger and Fear, as well as manifestations of an open and extraverted personality were designed specifically for this experiment and then adapted to the physical platform: a Sony entertainment robotic dog, AIBO ERS-210A. The choice of the robot platform was determined by its safety around humans, and the variety of expressive features it possesses, such as variable gaits, movable mouth, ears, tail, and LED display.

4.1 IMPLEMENTATION

4.1.1 BASIC FEATURES

The high-level controller was implemented in *MissionLab,* a robotics software toolkit [157], and the lower-level implementation was done in OPEN-R SDK-1.1.3, an open-source programming environment provided by Sony. When a user uttered a command, it was passed up to a Finite State Acceptor (FSA) in *MissionLab* by an administrator's key press, and then the processed command was passed on to the low-level controller on the robot.

There were a total of seven commands available: “Stop”, “Go Play”, “Follow the Ball”, “Kick the Ball”, “Follow Me”, “Come to Me”, and “Sic’ em”.

- “Stop” command stops the robot in a suspended state, ready to continue at any moment.

- “Go Play” makes the robot roam around in random directions and perform random stops (no additional user interaction is required for the robot to wander around beyond issuing the command once).
- In “Follow the Ball” mode, the robot looks for a pink ball by walking around in a circle, and moves towards the ball once it is detected; if the ball is lost, looking for the ball is resumed. Color recognition was used to detect the ball.
- “Kick the Ball” is similar to “Follow the Ball” except for when the robot is close enough to the ball to kick it, it performs a kicking motion.
- “Follow Me” command was identical to “Follow the Ball” command, but with a bouquet of artificial flowers used as a prop to follow instead of the ball.
- “Come to Me” is similar to the “Follow Me” command, but the robot stops if it comes within a certain distance to the prop.
- In “Sic’ em” command, the robot moves towards the “intruder robot” (ActiveMedia Amigobot) and stops next to it.

This basic set of behaviors was used in the Non-emotional condition, and augmented in the Emotional version via a variety of gaits, movement of ears and tail, LED changes, and minor behavior sequences, as described below.

4.1.2 PERSONALITY

Although a number of studies have established a link between non-verbal behavior and personality judgments [158, 159], identifying the specific behaviors characteristic to certain personality dimensions has proven to be much harder. Extraversion is the most studied personality trait in this respect, and was reported to positively correlate with an expressive, animated, and expansive behavioral style [158].

To separate personality from emotion, the encoding of personality was used only in the “Go Play” command, and the encoding of emotion was used in all other commands in this exploratory study. The personality was determined for the most part by the Extraversion dimension, as well as other traits (for some of the parameters). The parameters modified for the “Go Play” command were as follows: the proportion of time the robot wagged its tail was directly related to its level of Extraversion and Agreeableness; the probability of the robot changing gaits (Slow, Normal, Fast or Crawl) was directly proportional to its level of Extraversion and Openness, and the proportion of time the robot spent walking vs. stopping was inversely proportional to its level of Extraversion, and directly to its level of Neuroticism. Our intention was to present an energetic, friendly, and curious robot, and therefore the levels of Extraversion, Agreeableness and Openness were set to high, and the level of Neuroticism to low. The robot also turned its head as it was walking, which contributed to the display of energy and curiosity.

4.1.3 EMOTION

A number of sources were used to encode the display of emotions in the robotic dog. We conducted two informal surveys (11 and 21 people each) to find out laymen perceptions of dog emotions; consulted dog behavior literature [160, 161], and used commonsense to adapt the findings to fit within the technical limitations of the platform.

The expressive features used were as follows: three gaits (Normal, Fast – resembled slight jumping, and Crawling - somewhat similar to invitation to play/prowling behavior); three tail positions (up, flat, down); tail wagging; two ear positions (up and flat); and red illumination of LED screen.

The emotion expressed was determined by a combination of command type and presence of the command object (stimulus), except for the Sic’ em command, where the distance to the intruder robot also played a role. In particular, the following emotions were encoded:

- Interest – during the Ball commands, the dog used the crawling gait; when the pink ball was detected, the ears and tail went up. There were two slight variations on this emotion; they were expressed via different tail movements:
    - Alert – the tail was up when the pink ball was detected;
    - Friendly – the tail was wagging when the pink ball was detected.
- Joy – during the Follow Me command after the flowers were detected the dog used the Fast walk, and the ears went up. There were two slight variations on this emotion; they were expressed via different tail movements:
    - Active - the tail was up;
    - Friendly - the tail was wagging. When the robot was sufficiently close to the flowers, the tail was wagging faster.
- Anger and fear – during the Sic' em command, the robot used the Fast walk, the ears were flat, the tail was up and the LED screen was red until the dog robot got close to the intruder robot. After that, the red light went out, the tail and the head went down, and the robot backed up using the Crawling gait. A snapshot of the robotic dog "scaring off" the "intruder" by displaying "anger" is presented in **Figure 23**.

As different gaits were used in the Emotional and Non-emotional conditions, there were slight differences in performance: it was easier to kick the ball in the Emotional condition, and the Come To Me command was performed better in the Non-emotional condition. Also, the head following after object behavior was perceived by some subjects as emotional, although it was not intentionally encoded as such.

Figure 23: AIBO scares off the Amigobot by displaying “anger”

4.2 STUDY DESIGN AND ANALYSIS

The goal of the study two-fold: 1) to find out whether the presence of emotions and personality would increase the perceived ease of use and pleasantness of interaction, and 2) to identify whether the current implementation was sufficient to differentiate between a robot with emotions and personality from one without them. A longitudinal study allowed us to observe the human-robot interaction beyond a single short session, thus letting the “novelty” of a human subject interacting with a new robot to wear off.

4.2.1 EXPERIMENT DESIGN AND HYPOTHESES

The study was set up within a “robot as pet and personal protector” scenario, allowing for the exploration of relevant phenomena in a relatively constrained domain. During each session, the participants were requested to interact with the robot by asking it to perform certain tasks, with a new task introduced at each of the first three sessions. The participants were also encouraged to interact with the dog by petting it, playing with it, addressing it, and otherwise engaging with it if they so chose.

The study followed 1-factor independent design with two conditions: Non-Emotional and Emotional. In the Emotional condition, the robot’s basic set of behaviors was augmented with a display of emotions and personality via head, ear and tail position, a variety of gaits, and LED display (as described earlier in the implementation section), whereas in the Non-Emotional condition, the basic set of behaviors was left intact.

The following form the hypotheses of this experimental study:

1. *Hypothesis 1*: The display of emotions and personality will increase the perceived ease of use of autonomous robots;
2. *Hypothesis 2*: The display of emotions and personality will increase the pleasantness of interaction with autonomous robots and will generate greater attachment to them;
3. *Hypothesis 3*: The expression of emotions and personality will be more recognizable in the Emotional condition;
4. *Hypothesis 4:* The display of emotions and personality will result in higher Positive Mood and lower Negative Mood in the participants.

4.2.2 EXPERIMENT SETUP AND PROCEDURE

The study took place in a small quiet office. A Dell Latitude laptop was used to send subject-given commands to the AIBO wirelessly, and a Dell Precision 610 desktop was used to teleoperate the intruder Amigobot via RF. A green carpet was used to specify the borders within which the robot was to be kept by the users; they had an option of themselves staying on or off the carpet while interacting with the robot, and on or off a wheeled office chair The video camera was positioned on a desk overlooking the carpet, and the entire interaction between the participants and the robot was captured on video tape within the bounds of the carpet (prior to the experiment, the participants signed a video release form, allowing the researchers the use of their video footage). **Figure 24** gives a general overview of the setup.

Figure 24: Experiment Setup: the user is watching AIBO perform the “Kick the Ball” command on the green carpet.

In this longitudinal study, the subjects participated in four 20-60 minute interaction sessions. There were at least 3, but no more than 7 days between the sessions, for a total span of up to a month. Session duration included filling out questionnaires and the interaction with the dog per se, and depended on the willingness of the participants to interact with the robot. The overall goal behind the interaction was entertainment; however, some structure was provided, to both ensure that the participants could observe the entire range of affective expressions, and to make the sessions comparable across the participants.

The sessions resembled a series of dog-training exercises combined with some voluntary free play. Each session thus provided semi-structured interaction between participants and the robot, where new tasks for participants to perform were introduced at each session but the last (which was cumulative). Such semi-structured sessions allowed more control over participant’s actions, and thus made the conditions the same in everything but the robot’s affective behavior. In the affective condition, each task was associated with a different robotic emotion (or a variation of it), and with each session participants could observe a wider range of affect exhibited by the dog. At the end of each session the participants were also given a chance of interacting with the robot as they wished, using any commands they had trained the robot on. The combination of structured tasks (teaching the robot performing specific commands) and unstructured "play" time insured that the participants spent some minimum time interacting with the robot, but were not restricted to that time. Limiting the number of commands "taught" to the robot prevented the sessions from being overwhelming and overly long. A schematic view of the sessions' overall structure is provided in **Figure 25**.

Figure 25: A Schematic View of the Sessions

At the beginning of the first session the participants were asked to sign a consent form (*Appendix A*), and fill out a demographics questionnaire (*Appendix C*) and a personality questionnaire (*Appendix E*); then they were introduced to the robot and told that they could make the robot stop or "play" on its own, without requiring user interaction ("Stop" and "Go Play" commands, respectively). After that they were taught two more commands they could give the robotic dog: "follow the ball" and "kick the ball". Each command was first introduced, and then the participants were asked to repeat it three more times. These commands were to be separated by either "Stop" or "Go Play" commands. After the command that was introduced second ("kick the ball") was repeated three times, the subjects had an option of either continuing to interact with the robot, or completing the experiment by filling out the mood questionnaire (this questionnaire was filled out at the end of each of the four sessions; *Appendix G*).

In the second session, two new commands ("follow me" and "come to me") were introduced in the same manner as those in the first session. In **Figure 26**, you can see a subject getting the AIBO to perform the "Come to Me" command (Emotional Condition). After both of the commands were repeated three times, the subjects were asked to interact with the dog for at least five more minutes practicing the commands from this and the previous sessions. After this the participants, again, had an option of continuing the interaction or completing the session.

Figure 26: AIBO is performing the "Come to me" command in the Emotional Condition.

During the third session an "intruder robot" was introduced (another small robot, an Active Media robot Amigobot that was used for the purpose of testing the robot's role as a protector). As the Amigobot was teleoperated onto the carpet, the subjects were instructed to hide their props (the pink ball and the flowers) and give the AIBO a "Sic' em" command. After the robotic dog successfully performed the "Sic' em" command, the Amigobot was guided back off the carpet. Subsequent to the initial introduction of this command, the administrator brought out the "intruder robot" three more times, every four minutes. The participants were asked to interact with the dog using any of the commands they knew while waiting for the intruder robot to appear. After the last appearance of the Amigobot, the subjects could either continue the interaction, or complete the session.

Finally, during the last session the participants were asked to interact with the robot for at least fifteen minutes using any of the commands they knew in any order. The "intruder robot" was brought in from time to time (approximately every 3 to 4 minutes) throughout this session. At the end of the fifteen minutes the users could either continue the interaction, or proceed to fill out the rest of the questionnaires. The questionnaires were the mood questionnaire (*Appendix G*), the personality questionnaire regarding the robotic dog (*Appendix F*), and the post questionnaire (*Appendix D*). After all the questionnaires were filled out, some of the participants were given a choice of briefly interacting with the robot in the opposite condition than the one they were exposed to for all prior sessions (e.g., emotional instead of non-emotional). The administrator script can be found in *Appendix B*.

4.2.3 MEASURES

Evaluation was performed using both self-reports (questionnaires) and observation (videotapes analysis) methods with respect to the aforementioned study hypotheses.

The post questionnaire (*Appendix D*) was designed to assess hypotheses 1-3. It consisted of six 5-point Likert scale questions with three subquestions, with "Strongly

Agree" anchored at 5, and "Strongly Disagree" anchored at 1. The questions were as follows:

1. *It was easy to get the robotic dog perform the commands;*
2. *It was easy to understand whether the robot dog was performing the command or not;*
3. *The robotic dog showed emotional expressions;*
4. *The robotic dog had a personality of its own;*
5. *With every session, I was getting more attached to the dog;*
6. *Overall, I enjoyed the interaction with the robotic dog.*

If the participants answered "Agree" or "Strongly Agree" to question 3 or 4, they were also asked to answer questions 3a,b and 4a, respectively. The subquestions were as follows:

3a. Emotional expressions exhibited by the dog made the interaction more enjoyable;

3b. Emotional expressions exhibited by the dog made the interaction easier;

4a. I enjoyed interacting with the robot, partly because it possessed some personality.

Questions 1, 2 and 3b were used as measures for Hypothesis 1; questions 3a, 4a, 5, and 6 were used as measures for Hypothesis 2; and questions 3 and 4 served as measure for Hypothesis 3. In addition, at the end of the questionnaire the participants were asked a series of open-ended, free-response questions (**Figure 27**). The answers to these questions provided a less structured, qualitative approach of obtaining data.

Please use the space below (attach additional sheets if needed) to describe your interactions with the robotic dog. Specifically, did the dog seem to have a personality? If so, what kind of personality? Also, describe any emotional states that you think the dog exhibited during your interaction. Please describe your own state during the interaction: e.g., entertained, bored, curious, etc. Did your attitude change to the robotic dog throughout the sessions? How? Finally, would you prefer robots that interact with humans to express some emotion and personality? Why?

Figure 27: Free-response Questions from the Post Questionnaire

To analyze Hypothesis 4, separate Positive and Negative scores from the PANAS-T (Positive/Negative Emotionality Schedule, or "mood") questionnaire [162] were calculated for each session, as well as averaged across the four sessions.

Finally, the participants were also asked to fill out two personality questionnaires: the one regarding their personality was filled out at the beginning of the first session, and the one regarding the robotic dog's personality at the end of the last session. Both were based on a brief version of Goldberg's Unipolar Big-Five Markers (personality questionnaire) [163]; the dimensions assessed were Extraversion, Agreeableness, Conscientiousness, Emotional Stability (Neuroticism) and Openness to Experience/Intellect. Pearson's Correlation analysis was conducted on the participants' and the robot's personality to identify whether the subjects projected their own personality on the robot, but none of the dimensions had significant correlations.

4.2.4 PARTICIPANTS

A total of 20 people participated in the study: 10 males and 10 females, distributed equally between the two conditions. The subjects were recruited via flyers posted on and around the Georgia Institute of Technology campus, and they varied widely in the demographics according to age (from between 20 and 30 to over 50 years old), their educational level and backgrounds (from High School diploma to working on a Ph.D., with majority having either a Bachelor's or Master's degrees), and computer experience. Most of the participants had owned pets at some point in their lives (18 out of 20), and had either no or very limited robot interaction experience (only 2 out of 20 had interacted with mobile or entertainment robots prior to the study).

4.2.5 POST QUESTIONNAIRE ANALYSIS AND RESULTS

1-tailed Independent Samples T-tests were conducted on all the measures, and Pearson's correlations were computed where applicable.

4.2.5.1 Hypothesis 1: Perceived Ease of Use

The result of the analysis of question 1 (*"Easy to get the robot perform the commands"*) was statistically significant ($M_{non\text{-}emotional}$=3.7, $M_{emotional}$=4.5, F=0.02, p<0.004). The result is presented visually in **Figure 28**.

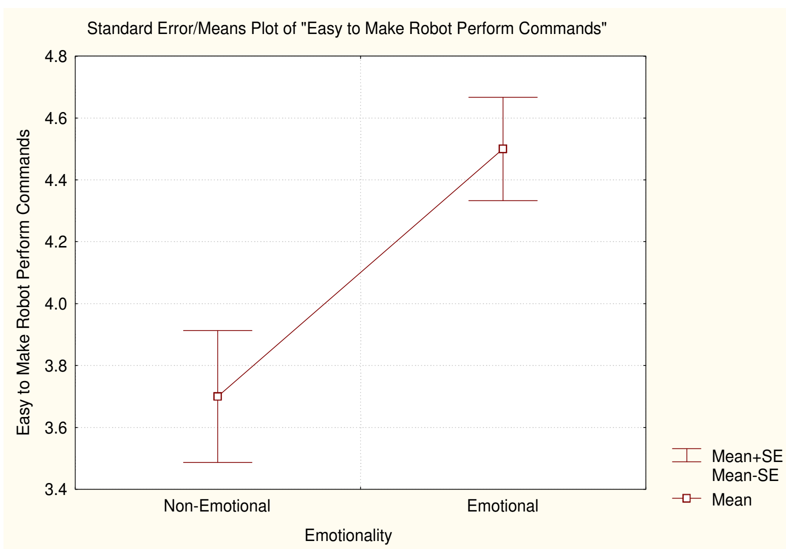


Figure 28: Standard Error/Means Plot for "Ease of Performing Command"; it was easier to make the robotic dog perform the commands in the Emotional condition

Although the analysis shows that the participants found it easier to get the robotic dog to perform the commands in the Emotional condition, this result is somewhat confounded by the slight performance differences for two of the commands. In particular, due to a difference in gaits, it was more difficult for the robot in the non-emotional condition to kick the ball, and vice versa, due to more frequent head movements performed by the robot in the emotional condition, it had more difficulty in accurately tracking the flowers during "Follow Me" command. Therefore the result could potentially be due to two factors: the performance difference and the emotionality; further study would be needed to disambiguate it. Special care was taken in the subsequent HRI experiments to ensure the same task performance in all conditions.

There was no significant effect of Emotionality on the answers to question 2 (*"Easy to understand if the robot was performing the command")*. As for question 3b (*"Emotional expressions made the interaction easier"*), the average answer was 3.48 (between Neutral and Agree), suggesting that those who thought that the robot displayed emotional expressions (5 out of 10 in the Non-Emotional condition, and 8 out of 10 in the Emotional condition), considered emotional expression somewhat helpful in making the interaction with the robot easier.

4.2.5.2 Hypothesis 2: Pleasantness of Interaction and Level of Attachment

The analysis of the answers to questions 5 (*"Getting more attached"*) and 6 (*"Enjoyed the interaction overall"*) did not show any significant difference between the two conditions with respect to the overall pleasantness of interaction and the degree of attachment to the robot. However, those who believed that the robot displayed emotions and/or personality (6 out of 10 in each condition), also believed that these features made their interaction with the robotic dog more pleasant: the average answer for question 3a (*"Emotional expressions made the interaction more enjoyable"*) was 4.46, and for question 4a (*"Robot's personality influenced the level of enjoyment"*) was 4.25, both between Agree and Strongly Agree. These results are shown graphically in **Figure 29**.

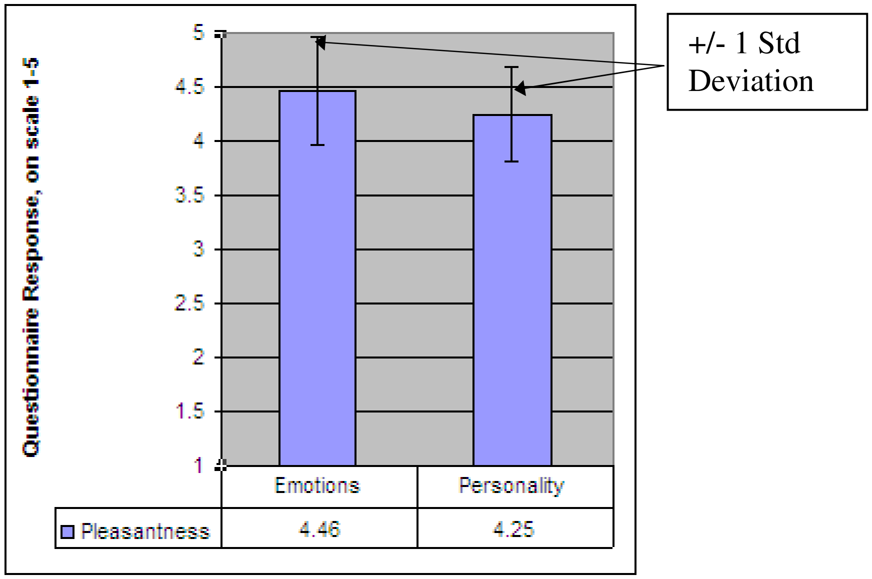


Figure 29: Average of the answers to questions 3a and 4a: the subjects thought the perceived robotic emotions or personality made their interaction more enjoyable.

There were also a number of significant correlations between questionnaire responses regarding the pleasantness of the interaction:

1. The response to question 4 (*"Robotic dog had a personality"*) had a strong positive correlation (r = .66, p=0.02) with the response to question 5 (*"Getting more attached"*);

2. The response to question 3a (*"Emotional expressions made the interaction more enjoyable"*) had strong positive correlations with responses to questions 5 (r=617, p=.025) and 6 (r=.749, p=.003).

Overall, the participants mostly enjoyed their interaction with the robot, as evidenced by the average score of 4.27 out of 5 in answer to question 6 (*"Enjoyed the interaction overall"*), and the perceptions of the robot's emotionality and personality seem to make the interaction more enjoyable and result in greater attachment.

4.2.5.3 Hypothesis 3: Recognition of Emotions and Personality

There was no significant difference between the two conditions regarding perceived emotional ($M_{non\text{-}emotional}$=2.7, $M_{emotional}$=3.6, F=.693, p<0.088), and personality ($M_{non\text{-}emotional}$=3.1, $M_{emotional}$=3.56, F=.1, p<0.324) display (questions 3 and 4). However, a 2-factor ANOVA on Gender and Emotionality resulted in a significant main effect of Gender on the answer to question 3: display of emotions (M_{female}=3.8, M_{male}=2.5, F =4.829, p<0.043). The following graph presents this result (**Figure 30**):

Figure 30: Standard Error/Means Plot for "Displayed Emotions" by Gender: Women perceived the robot as more emotional than men, regardless of the condition.

In a related study, Yan et al [60] manipulated the encoded Introversion and Extraversion dimensions of personality successfully, suggesting that people do pick up nonverbal and verbal personality cues in robots, even though this was not confirmed in our study. One of the reasons we did not see any difference in perceived personality between the two conditions in our study may lie in the fact that the personality was encoded only in one task, "Go Play", and the subjects did not have sufficient exposure to its display. Another reason could be that our encoding was broader and less specific than in the aforementioned study, and thus did not produce the expected effect.

4.2.5.4 Hypothesis 4: Changes in Positive and Negative Moods

Although there was no significant result of Emotionality on the Positive Mood, the Negative Mood was significantly lower in the Emotional condition ($M_{non\text{-}emotional}$=13.9, $M_{emotional}$=12.125, F=6.462, p<0.048). See **Figure 31** for the plot.

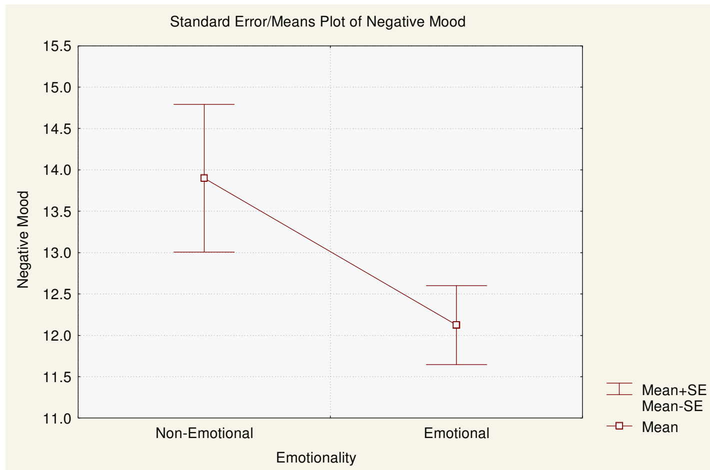


Figure 31: Standard Error/Means Plot for Negative Mood: on average, the subjects experience less negative mood in the Emotional Condition

In the Positive/Negative Emotionality measure (PANAS-T, [162]), negative mood refers to the extent to which an individual is presently upset or distressed, and positive mood refers to one's current level of pleasure and enthusiasm. Therefore, a lower negative mood experienced by participants signifies lower levels of distress and frustration, and in turns signals a more pleasant interaction. Additionally, a significant positive correlation (r=.598, p=.007) between average Positive Mood and the response to question 4 (*"Robot displayed personality"*) was observed, thus providing a link between perceived robotic personality and users' improved mood.

4.2.5.5 Other Observations

We have also observed that the subjects in the Emotional condition rated the robotic dog higher on the dimensions of Conscientiousness and Openness, and in the 1-tailed independent samples T-test this difference was statistically significant (p<.034, and

$p<.026$, respectively). The last finding, however, may also be convoluted by the aforementioned difference in the performance. In the future, extensive testing will be done to eliminate potential task performance differences.

4.2.5.6 Summary and Discussion

To summarize, two of the hypotheses were confirmed in this study, and a number of interesting and encouraging observations were made:

1. The participants found it **easier** to get the robot in the Emotional condition to perform commands;
2. The **Negative Mood** reported by the participants in the Emotional condition was **lower** than by those in the Non-emotional condition; a lower negative mood signifies lower levels of distress and frustration, suggesting that affective behavior contributes to the quality of interaction.
3. Those participants who believed that the robot displayed emotions and/or personality also believed that these features made their interaction more pleasant. This is encouraging, as it suggests that people value expression of emotion and personality in their interaction with an autonomous entertainment robot.
4. Women were found to be more attuned to emotional expressions and more ready to attribute emotions to the robot than men, which should be taken into consideration for systems adapted to groups with gender-biased compositions.

There are a number of reasons why no differences were observed in the post questionnaire in the area of pleasantness of interaction and perception of emotions and personality. First, the sample size, due to the technical complexity of the study, was rather small; in the future studies, a larger number of participants per condition will be used. Second, people without extensive robotics experience seem to have substantial preconceptions of robots, and may project their own affective state onto the robot they interact with. Finally, the physical platform, designed for entertainment, may be a more decisive factor than the actual differences in the robot behavior, and may have influenced the perceptions of the robot as that of a playful, expressive toy dog.

4.2.6 FREE-RESPONSE QUESTIONNAIRE ANALYSIS AND RESULTS

As a reminder, the free-response portion of the questionnaire went as follows:

> Please use the space below (attach additional sheets if needed) to describe your interactions with the robotic dog. Specifically, did the dog seem to have a personality? If so, what kind of personality? Also, describe any emotional states that you think the dog exhibited during your interaction. Please describe your own state during the interaction: e.g., entertained, bored, curious, etc. Did your attitude change to the robotic dog throughout the sessions? How? Finally, would you prefer robots that interact with humans to express some emotion and personality? Why?

The summary of the participants' responses to these questions is provided in *Appendix H.* In order to better understand the opinions given by the participants by means of the free-response portion of the post questionnaire and to obtain quantitative data, two independent coders (neither of whom was the experimenter) were asked to rate the participants' responses according to a number of scales. These scales correspond to the hypotheses presented in subsection 4.2.1, *Experiment Design and Hypotheses*. In particular, questions 1 and 2 measure how well participants could recognize emotions and personality in the robot, and correspond to hypothesis 3 regarding participants' perception of emotions and personality in the robot; and questions 3-6 measure the participants' state throughout the interaction and correspond to hypothesis 2 regarding pleasantness of interaction. These scales are anchored at "*Very Low*" at 1 and "*Very High*" at 7.

1. On a scale from 1 to 7, please rate the level to which the participant perceived emotions and/or personality in the robot;
2. On a scale from 1 to 7, please rate the level of detail in the descriptions of emotions and personality are;
3. On a scale from 1 to 7, please rate the participant's boredom during the interaction;
4. On a scale from 1 to 7, please rate the participant's enjoyment during the interaction;
5. On a scale from 1 to 7, please rate the participant's frustration during the interaction;
6. On a scale from 1 to 7, please rate the participant's contentment during the interaction.

In addition, to assess whether the participants' attitude towards the robot changed throughout the study, the coders were also asked to provide a rating for Question 7:

7. On a scale from 1 to 7, please rate the change in participant's attitude towards the robot throughout the sessions.

This scale was anchored at "*Overwhelmingly to the Worse*" at 1 and "*Overwhelmingly to the Better*" at 7. The instructions given out to the coders can be found in Appendix I.

4.2.6.1 Analysis and Discussion

The scores obtained from both coders were averaged, and the values marked as "Not Reported" by either coder were considered missing data and excluded from the analysis. A one-tailed T-test was conducted on the scores for all the questions but two: questions 5 (measuring level of frustration) and 7 (level of change) were excluded from the analysis due to a high percentage of "Not Reported" scores (50% and 30%, respectively). No statistically significant differences were found in the scores of the analyzed questions between the emotional and the non-emotional conditions; this confirms earlier findings from the Likert-style portion of the questionnaire regarding perceived emotions and personality in the robot (subsection 4.2.5.3) and the level of enjoyment (subsection 4.2.5.2). However, given the exploratory nature of the study, one interesting observation can be made: across both conditions, the average score for question 1 (level to which participants perceived emotions and/or personality in the robot) was 4.3, between medium and somewhat high (see Appendix I). This suggests that people were prone to see some affect in the robot, whether it was encoded or not and in line with the earlier discussion (subsection 4.2.5.6) that the physical platform may have biased the perception of affective behavior. It was also interesting to see that overall the participants enjoyed their extended (over 2 hours combined in some cases) interaction with the robot to a medium/somewhat high level (score of 4.48) and were only slightly bored (to a low/somewhat low level with a score of 2.56); this can be seen graphically in **Figure 32**.

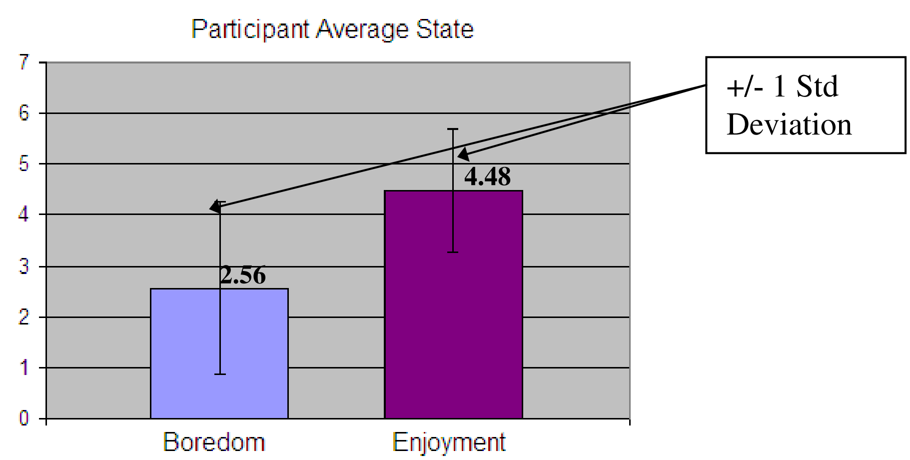


Figure 32: Average score of participants' boredom and enjoyment, surmised from the free-response questionnaire

In addition to analyzing the data provided by the coders, participants' responses to whether they would prefer robots that interact with humans to express some emotion and personality were also counted. The responses to this question were rather straightforward: either yes or no, or both yes and no; in case of the ambivalent responses, a justification was given; two of the participants didn't answer the question. The majority of participants (15 out of 20) would prefer robots that interact with people to **express some emotions and/or personality**, at least in certain cases. Their responses are summarized in **Figure 33**. Among those who were ambivalent (two in the non-emotional condition, where no robotic affect was expressed, and 4 in the Emotional condition), four reported that emotions and personality would be desirable for some purposes and applications, but not for others; one was concerned with people forming attachment to emotional robots; and yet another one said that better implementation of such features would be useful (as compared to Microsoft's Paperclip). Finally, none of those in the emotional condition said a definite "no" to robotic affect.

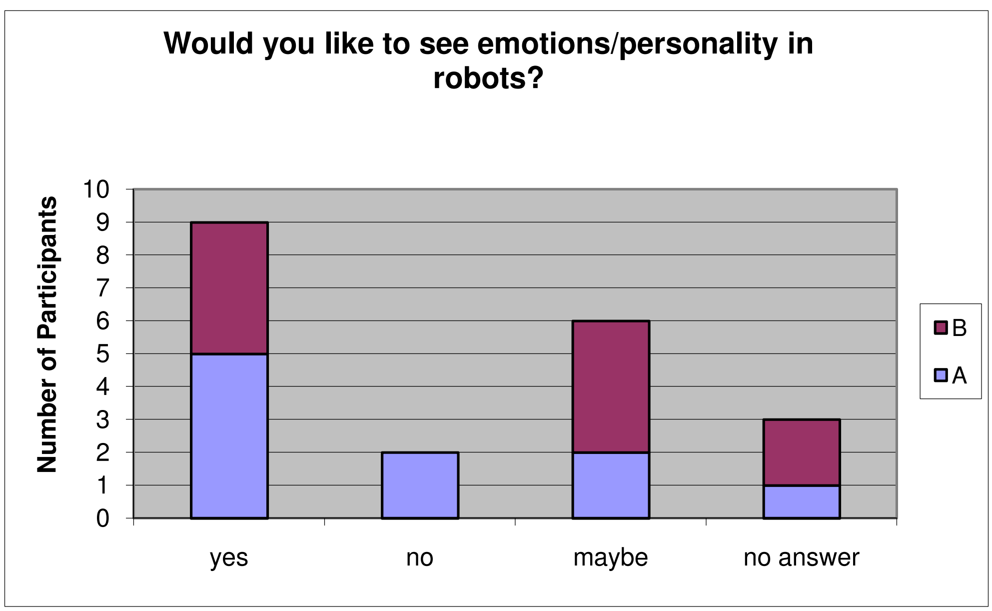


Figure 33: Response to the Question on whether Participants Prefer Robots to Express Emotions and/or Personality

Finally, the reasons given for including personality and emotions in robots can be subdivided into three somewhat overlapping categories: 1) for certain applications for which they contribute to the overall goal; 2) to make the interaction more pleasant, enjoyable, personal and creative; 3) because humans use and need emotions in their interaction. There were two responses that did not quite fit within these categories: one person suggested that emotional response in robots would improve usability and communication of commands (thus contributing to the ease of interaction); and the other one touched the issue of trust: people would be less intimidated and more comfortable dealing with robots. The primary justification for ***not*** using affect in robots was using it for "inappropriate" tasks, such as strictly functional or mundane, while others (two responders) were more concerned with ethical issues, such as causing attachment in humans and the fact that robots are not human.

4.2.6.2 Summary

Overall, the results of the analysis of the free-response portion of the questionnaire were in line with the results obtained in the analysis of the Likert-scale questions. Although there were no statistically significant differences in the levels of enjoyment, boredom and contentment between the two conditions, and in the reported expressions of emotions and personality, it was found that the subjects expected manifestation of affect in a robotic dog:

8. The participants were prone to attribute affect to the robot, whether it was intentionally encoded or not;
9. The majority (15 out 20 participants) would prefer robots that interact with people to have emotions and/or personality, at least for certain applications.

4.2.7 VIDEO ANALYSIS

As video analysis by independent coders is both time- and resource-consuming, a partial video coding was done by the experimenter to address the feasibility and usefulness of engaging at least 2 independent coders to confirm or refute any significant findings (the entire set of recorded sessions would be coded independently by such individuals). The video sessions of 12 out of 20 participants (equally distributed between the conditions and randomly selected per condition) were coded for a number of measures presented below, and the obtained data were collected and partially analyzed.

4.2.7.1 Coding Measures

The measures for video coding were selected prior to the experiment. The main focus was on assessing the level of interaction with the robotic dog, which could be indicative of the quality of interaction, in particular, pleasantness (Hypothesis 2). Unfortunately, the facial expressions of the participants were not recorded; therefore we could not assess the level of enjoyment or interest directly. The following measures were used during the coding of the video sessions:

1. *Total Time* was recorded for each session. It excluded the time during which the robot may have been unresponsive or down (this happened in one out of 48 coded sessions). Although not a direct measure of interactivity, it could be indicative of how long the subjects were willing to interact with the robot. However, this measure could be only used as supportive, as it would also reflect how hard it was to perform the commands: the harder it was to get the robot to perform commands, the longer the participants would take.

2. *Extra Time* was recorded for each session. At the end of each session, after all the mandatory tasks were completed, the participants were given an opportunity

to interact with the robot longer if they so desired. This was thought as a primary indicator of how much people enjoyed their interactions.

3. *Percentage of Time on the Rug*. The participants were instructed to keep the robotic dog on the provided green rug; at the beginning of the first section they were also encouraged to walk on the rug if they wanted. While selecting this measure, it was hypothesized that those who stayed on the rug for longer portion of a session were more interactive.

4. *Percentage of Time spent closer to the Level of the robot*. During the sessions, some of the participants were either sitting/squatting on the floor, bending down while standing, or bending down in the chair provided off the rug. It was hypothesized that those who were purposefully closer to the level of the dog were more interactive.

5. *He vs. it*. Some participants referred to the dog as he, and some as it. It was hypothesized that those who called the dog "he" viewed it as more animate and life-like.

6. *Positive Speech Utterances*. Those included praise/approval ("good boy", "good job", "there you go", and more generic "Yeay!", "Wow!", etc.), expressions of concern ("Oooh", "ouch!", etc.), and talking to the robotic dog directly, but perhaps in more neutral terms (this shows attempts at reciprocity [70]). These were hypothesized to be an indicator of participants' enjoyment, perception of the robot as life-like, as well as their general attitude towards the robot; in pet interactions, such remarks show approval and affection.

7. *Negative Speech Utterances*. Those would include discouragement/punishment statements, such as “bad dog”, etc. In the 48 coded sessions no such negative statements were made by the participants.

8. *Positive Actions*. A number of participants petted the dog, pointed in the direction of the prop as if encouraging the dog to follow, leaned close to the robot, and moved the “flowers” prop back and forth in front of the robot to watch it follow with its head (some seemed to enjoy this a lot). These actions indicate affection and attempts at reciprocity as discussed in Dautenhahn et al. [70] .

9. *Negative Actions*. On very few occasions participants would make motions as if they’d hit the robot, either with their foot or the “flowers” prop; would pick up the robot into the air, turn it over and watch it move its legs; watch the robot trample the flower prop or walk into an obstacle. These occasions were classified as mistreatment [70] and were viewed as a sign of general dissatisfaction and treating the robot as fully inanimate.

10. *Duration and number of commands.* Although this measure would not assess the interactivity and enjoyment directly, it would allow identification of particular commands (if any) that participants enjoyed the most, and whether they differed between the conditions. If any differences were found, we could concentrate on specific commands to identify the reasons for the differences.

4.2.7.2 Preliminary Analysis and Discussion

Means and Standard Deviations were calculated for each measure, except for those dealing with speech (*positive utterances* and *he vs. it* measure), and *positive and negative actions*. The latter measures were fairly infrequent, and to an extent subjective; they seemed to depend greatly on individuals. For example, some participants could be using a lot of speech and a fair number of actions, whereas others would be mostly silent throughout the sessions, even though all the participants were equally encouraged to articulate their feelings prior to the experiment. In addition, the measures of *Time on Rug* and *Time Level Down* were fairly subjective, as a judgment call had to be made for borderline cases (e.g., one foot on the rug; squatting, but not leaning towards the dog or interacting with it, etc.)

Unfortunately, the initial results were not encouraging: the obtained data for 48 sessions were not normally distributed, and had very high standard deviations (in some cases they exceeded the means). This variation in results can have a number of possible explanations, or more likely, a combination thereof: 1) individual differences in the participants were on the extreme side, with some people scoring very high and some very low; 2) the chosen measures were not fully adequate; 3) the number of sessions coded was too low to provide adequate results; 4) insufficient pilot testing, in particular, with regards to metrics. For the time being, due to the time-consuming nature of the task and high data variation in the preliminary analysis, it was not feasible to obtain the services of two independent coders to complete the video analysis of the study.

4.3 SUMMARY AND LESSONS LEARNED

To summarize, the hypotheses, regarding the ease of interaction and effects on participant's mood, were confirmed in this study, and a number of interesting and encouraging observations were made:

1. The participants found it **easier** to get the robot in the Emotional condition to perform commands.
2. The **Negative Mood** reported by the participants in the Emotional condition was **lower** than by those in the Non-emotional condition; a lower negative mood signifies lower levels of distress and frustration, suggesting that affective behavior contributes to the quality of interaction.

3. Those participants who believed that the robot displayed emotions and/or personality also believed that these features made their interaction more pleasant. This is encouraging, as it suggests that people value expression of emotion and personality in their interaction with an autonomous entertainment robot.

4. The participants were prone **to attribute affect** to the robot, whether it was intentionally encoded or not.

    - Women were found to be more attuned to emotional expressions and more ready to attribute emotions to the robot than men, which should be taken into consideration for systems adapted to groups with gender-biased compositions;

5. The majority (15 out 20 participants) stated that they would **prefer robots** that interact with people **to have emotions and/or personality**, at least for certain applications.

Most importantly, a number of lessons were learned for future experiment design with affective robots. Although they may seem obvious in retrospect, the aspects addressed below were not intuitive, especially given the dearth of work done in the area of assessing user perceptions of affective robots at the time. Some of the points listed below can be applicable to any experimental design, but are especially important for the domain of HRI. These lessons are discussed below, along with improvement suggestions for future HRI experiments.

1. **Physical platform**. It is fairly clear from the results that the choice of the physical platform can dominate user perception of the robot's personality and emotion, as well, perhaps, as their enjoyment level. Sony's AIBO was designed for

entertainment, and many users described it as “cute” and “puppy-like” before the actual interaction with the robot. Therefore it was easier to attribute personality and emotionality to the robotic dog from the onset, as the participants were primed by the robots’ shape and movements. For future studies, a more neutral platform should be chosen, and an extensive pilot study should be done to assess the recognition of affective robotic expressions by participants before the actual experiments are performed. It is also likely that with a physical platform that is on the other side of the spectrum – openly mechanical and non-interactive, an opposite effect may be observed: participants may be less likely to attribute any human/animal-like qualities to the robot, and enjoy the interaction less.

2. **Robot Performance Differences**. There were slight differences in performance of certain tasks (“Kick the Ball” and “Follow Me”), due to which some of the results regarding ease of use were convoluted. In particular, due to a difference in gaits, it was more difficult for the robot in the non-emotional condition to kick the ball, and vice versa, due to more head movements performed by the robot in the emotional condition, it had more difficulty in accurately tracking the flowers during “Follow Me” command. For future studies, special care should be taken to avoid such differences, and the robot’s movement and task performance should be tested and compared between the conditions prior to the pilot study.

3. **Strength of affect encoding**. This aspect refers to how different the conditions appear to the subjects, and is related to the chosen physical platform. In particular, in a neutral platform, even smaller programmed differences may produce a desired effect, whereas in a biased platform the differences may need to be more pronounced. Increasing the differences between conditions will also

improve the statistical power of the experiment [164], or the probability that the null hypothesis will be rejected when it is false. Again, extensive pilot studies should be performed to assess whether the strength of encoding was sufficient to produce the desired manipulation.

4. **Sample size.** A larger sample size would be needed to produce more statistically significant results, especially given that the differences between the conditions were perceived as small. Due to difficulties in conducting experiments with real robot platforms, small sample size is a common pitfall for HRI studies. For future studies, all efforts should be made to obtain a larger number of participants per condition.

5. **Within-subject vs. between-subject design**. At the end of the study, some of the participants were shown the opposite condition, and most of them at that point could identify the differences between the conditions, and point out the emotional expressions in the Emotional condition. It would seem, then, that structuring the study as a within-subject design (where the same subjects participate in both conditions) would bring out the differences between the conditions clearer. Within-subject design would also reduce variability due to individual differences, thus also increasing the power of the experiment while requiring fewer participants that between-study designs [164]. However, if the physical robot is the same, the subjects might get confused if they see the same robot behaving differently, and it also may not be appropriate for certain tasks.

6. **Individual differences.** Unlike the majority of experiments, where the user base is comprised of predominantly undergraduate psychology students (or often fellow computer scientists, in the area of HCI), the participant composition of this study was quite varied. If the sample size were substantially larger, this variety

would have been desirable; however, with a smaller sample size, it may have resulted in the observed high variation of subjects' interaction with the robot.

7. **Data Collection**. One aspect the data were missing were videos of participants' faces, which could provide information on smiles and other facial expressions indicative of positive or negative emotions. For any video analysis attempting to assess the levels of enjoyment and satisfaction it would be advisable to include faces in the recording.

8. **Metrics.** A special effort should be made to provide metrics that can indeed measure the effect of affective behavior on human-robot interaction. First of all, the proposed metrics should be tested in a pilot study and refined prior to the experiment proper. Whenever possible, an expert opinion should be obtained on any designed measures to ensure they are free of bias, and protocols for coding video recordings should be developed prior to the experiment and refined as a result of the pilot study.

The lessons learned as a result of this exploratory study were carefully examined and taken into consideration during the design and administration of the subsequent HRI experiments used to evaluate the *TAME* framework. In particular:

- An affect recognition survey was conducted prior to the experiments to increase the probability that the encoded affective expressions would be recognized by the subjects;
- Performance differences were brought to a minimum;
- Novel subjective metrics were developed and combined with compliance and task performance objective measures;

- The affective robotic expressions were exaggerated to increase the difference between conditions;
- Number of participants per condition was increased;
- Extensive pilot testing was performed.

5 SOFTWARE ARCHITECTURE AND IMPLEMENTATION[1]

In chapter 3, a computational theory and psychological foundations behind the *TAME* framework were presented, and chapter 4 described an exploratory study undertaken to inform the development process of the framework and assess people's attitude towards robotic affect. This chapter describes the software architecture and implementation of the computational theory, and, as such, it addresses the first research subquestion, *"How can traits, attitudes, moods and emotions be modeled computationally in a robotic system"*, by grounding the theory into a particular software system and implementation. Additionally, this chapter presents an online survey performed to evaluate the recognition of affective robotic expressions [165], as implemented on a humanoid robot. The survey provides an initial testing step, necessary to confirm that the designed affective behaviors are recognized as intended by independent participants and can, therefore, be used in subsequent HRI studies.

5.1 SOFTWARE ARCHITECTURE DESIGN

The software architecture incorporating the *TAME* framework was designed as a stand-alone process to achieve platform-independence and generalizability [166]. With an interface to connect to the system's *TAME Communication Manager* (to supply sensory data), and appropriate configuration files, this software can potentially be integrated into any robotic platform or an autonomous agent system without a substantial redesign. The architecture itself is fairly straightforward, and consists of: *TAME Manager* (the central module of the system), *TAME Communication Manager* (receives sensor data and passes the updated affective values to the robot), a module

[1] To a significant extent the software design and implementation of the framework was performed under a 2-year grant from Samsung Electronics Co., Ltd, with the generous help of Sunghyun Park, Hyunryong Jung and Chien-Ming Huang who were supported through this grant. The project resulted in a number of conference and journal publications [163-165], portions of which were used in the write up of this chapter.

for each of the affective components, and *Stimuli Interpreter* (processes incoming sensory input). In addition, 2 configuration files, *Affect Configuration and Stimuli Configuration,* specify: interdependencies between the affective components; bounds for affective variable parameters; and mapping between sensor data types and affect generation. These configuration files provide a way of adjusting affect generation parameters to particular platforms and environments without reprogramming, thus increasing software flexibility. Finally, *Case Library* stores a set of attitude cases for each user interacting with the system. **Figure 34** provides a high-level view of the *TAME* module.

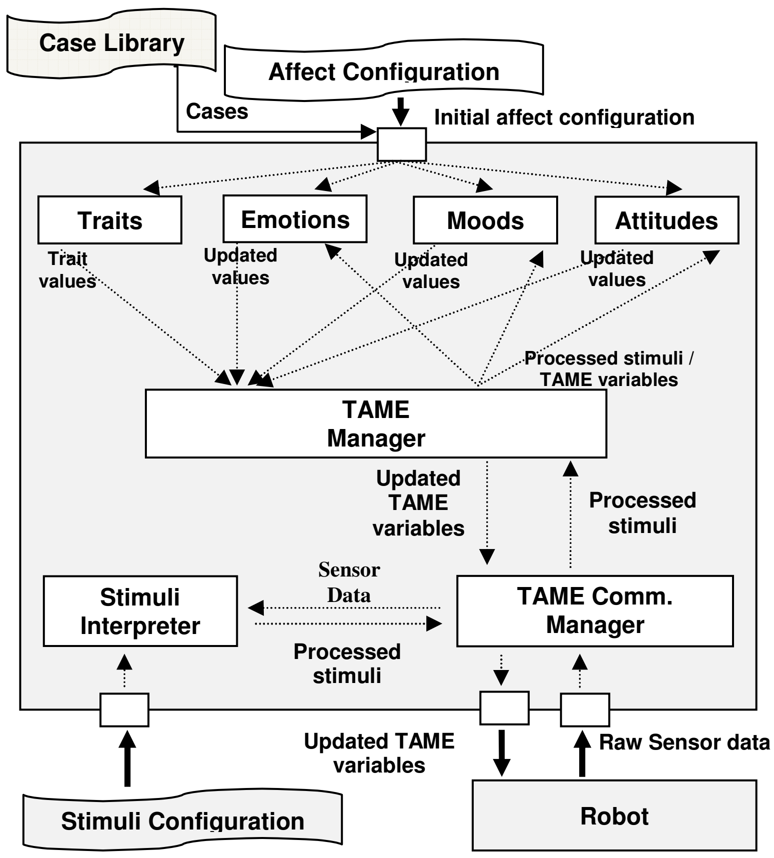


Figure 34: High-level View of the TAME Software Architecture.

5.1.1 AFFECTIVE COMPONENTS

These are comprised of four different affective components of the TAME framework (namely *Trait*, *Attitude*, *Mood*, and *Emotion*), and run as separate threads. Each module processes sensory data and internal information (current values of other affective components) received from *TAME Manager* and calculates the updated affective variables, passing them back to *TAME Manager*, and eventually to a robot or agent controller. In order to provide flexibility and adaptation to individual users and situations, each component is preloaded with some initial default values from the *Affect Configuration* file. An annotated sample *Affect Configuration* file can be found in Appendix J.

5.1.1.1 Trait Component

For the *Trait* component, a default value is specified in the *Affect Configuration* file (see Appendix J for a sample configuration) for each of the five personality dimensions: *Openness (O), Agreeableness (A)*, *Conscientiousness (C)*, *Extraversion (E)*, and *Neuroticism (N).* Psychological findings of the normal distribution of human personality scores from the Five-Factor model tests serve as a point of reference for selecting the trait intensity and range values, as described in subsection 3.2.2.1, *Grounding Trait Intensity in Psychological Data*. The means and Standard Deviations (SD) of these scores are presented in **Table 12:** Normal distribution statistics for FFM personality scores

Table 12: Normal distribution statistics for FFM personality scores

Personality Trait	Mean	Standard Deviation (SD)
Openness	110.6	17.3
Conscientiousness	123.1	17.6
Extroversion	109.4	18.4
Agreeableness	124.3	15.8
Neuroticism	79.1	21.2

For example, if the desired personality is highly agreeable and slightly neurotic, then a value of 148 (mean + 1.5 SD) can be assigned for Agreeableness, and a value of 89.7

(mean + 0.5 SD) for Neuroticism. **Table 13** presents the personality configuration for such a robot, where the rest of the traits are set as average (mean). The same personality configuration is presented graphically in **Figure 35:** A sample personality configuration: Agreeableness is moderately high, and Neuroticism is slightly elevated

Table 13: A sample personality configuration for a highly agreeable and slightly neurotic robot

Trait	Value
Openness	110.6
Conscientiousness	123.1
Extroversion	109.4
Agreeableness	148
Neuroticism	89.7

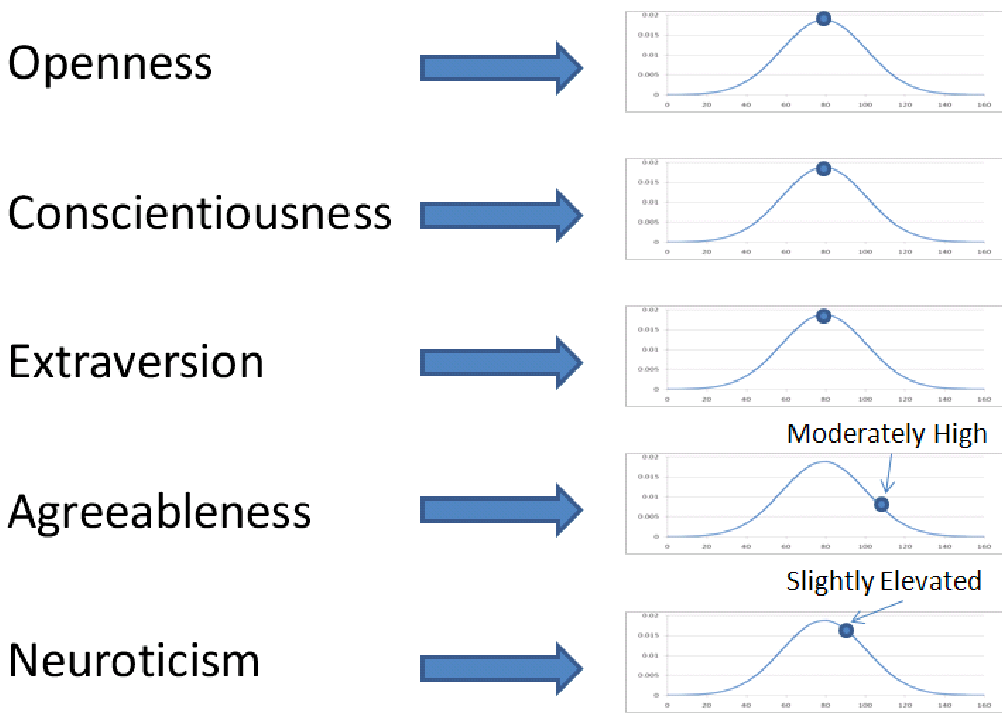


Figure 35: A sample personality configuration: Agreeableness is moderately high, and Neuroticism is slightly elevated

Once the trait values are specified, they become available to Trait component and to the *TAME* Manager; the latter in turn provides them whenever they are needed by other components. Trait values remain unchanged throughout execution since personality is generally regarded to be time-invariant. There is one exception to this, however: for computational simplicity, as moods affect the same behaviors as traits, instead of providing a direct influence on behavior, the moods temporarily bias trait values, which, in their turn, provide an adjustment to the underlying behaviors. This exception will be discussed in more detail in subsection 5.1.1.3, *Mood Component*.

5.1.1.2 Emotion Component

This component contains six primary emotions (*Fear (F), Anger (A), Disgust (D), Sadness (S), Interest (I) and Joy (J)*). From the *TAME Manager*, it receives emotion-specific stimuli strengths (calculated in *Stimuli Interpreter*) and current trait and mood values. The values of each emotion are generated continuously throughout the execution whenever emotion-eliciting stimuli strengths are relayed to the emotion component by the TAME Manager; the updated emotion intensities are then passed back to the TAME Manager.

At the start, the Emotion component is initialized with default settings from the Affect *Configuration* file (Appendix J). These settings are used during the execution for emotion generation and would differ depending on a particular task, mission or an environment (although an administrator/user can provide a set of generic defaults as well). They include:

- Upper and lower bounds for all emotion generation variables for each emotion: amplitude (controlling the peak of emotion), activation point (controlling emotion sensitivity is to an eliciting stimulus), and slope (controlling how fast emotion rises in response to a stimulus); see subsection 3.3.2.2.1, *Personality Influence on Base Emotion Generation* for details on these variables ;

- Emotion Decay Rate (subsection 3.3.2.3, *Emotion Decay and Filtering*);

- Filtering variables: Prior Weight and Current Weight for a weighted averaging filter (subsection 3.3.2.3, *Emotion Decay and Filtering*);
- Personality/emotion dependency matrix (a sample matrix is presented in **Table 14**). This matrix specifies how each trait influences each of the emotions: 1 stands for direct influence, -1 for inverse influence, and 0 for absence of influence. For example, the sample matrix in **Table 14** specifies that Agreeableness has an inverse influence on Anger generation, indicating that a robot high on agreeableness would experience anger to a lesser extent than its more disagreeable counterpart. Thus, for each different personality configuration, the emotions are generated using different values of emotion generation variables.

Table 14: Sample Personality/Emotion Matrix specifying the influence traits have on the emotion generation variables

TRAITS / **EMOTIONS**	**O** Openness	**A** Agreeableness	**C** Conscientiousness	**E** Extraversion	**N** Neuroticism
Interest	1	0	0	1	0
Joy	1	1	0	1	0
Fear	1	0	0	0	1
Anger	1	-1	0	0	1
Sadness	1	0	0	0	1
Disgust	1	0	0	0	1

Although it would be possible for an advanced user to select the settings in the *Affect Configuration* file to suit a particular task, in general, providing these defaults would be best left to the designer or administrator, as they would influence complex interactions within the module.

Throughout mission execution, whenever emotion-eliciting stimuli information is received by the TAME Manager, it sends an update request to the *Emotion* component, along with the stimulus strength data and current mood and trait values. Every time an update request is received by the Emotion component, the following steps are taken for each emotion:

1. Calculate emotion generation variables: amplitude, maximum slope and activation point;:
    - Calculate trait-based emotion generation variables: amplitude (eq. 7 and 8), slope (eq. 9 and 10) and activation point (eq. 11 and 12), as described in subsection 3.3.2.2.1, *Personality Influence on Base Emotion Generation*;
    - Calculate mood-based portion of the activation point variable, as described in subsection 3.3.2.2.2, *Mood Influence on Base Emotion Generation*, using eq. (13);
    - Calculate the overall activation point variable using equation (14), as described in subsection 3.3.2.2.3, *Combined Personality and Mood Influence;*
2. Generate base emotion intensity taking into consideration the stimulus strength and emotion generation variables (equation (6), subsection 3.3.2.2, *Emotion Generation*).
3. Update the emotion intensity in accordance with decay function and filtering, using equations (15) and (16), respectively, as described in subsection 3.3.2.3, *Emotion Decay and Filtering*;
4. Pass the emotion intensity values to TAME Manager after they have been updated.

The pseudocode for Emotion update can be found in Appendix L.1.

5.1.1.3 Mood Component

The *Mood* component maintains the levels of Positive and Negative moods based on the perceptual information regarding environmental and internal conditions obtained through the TAME Manager. The values of each mood are updated continuously throughout the execution whenever new environmental/internal condition levels are relayed to the *Mood* component by the *TAME Manager*; the updated levels are then passed back to the TAME Manager.

During the software design, a slight variation in how the base mood is calculated was made: a weighted average of all environmental and internal conditions which are of relevance to a particular mood is used (eq. 35) as opposed to a weighted summation of these influences (eq. 18, subsection 3.4.2.1, *Mood Generation*). This was primarily done for computational simplicity, and does not affect the overall design. In this case, the base mood level is calculated not within the Mood component itself, but is rather assigned the overall environmental/internal conditions level computed as a weighted average by the Stimuli Interpreter (subsection 5.1.3, *Stimuli Interpreter*). Once the base mood is once assigned, a sliding window averaging filter is applied to it, to smooth the effect of any sudden changes in conditions.

In addition to maintaining the mood levels, the Mood component also temporarily adjusts the trait intensities. Originally, the *Mood* component was designed to produce a direct, albeit small influence on behavioral parameters (subsection 3.4.3, *Influence on Behavior*), while traits specified the base parameters to be used for the entire mission. However, for computational simplicity, a deviation from the original framework design was made, without the loss of generality. In particular, given that both mood and traits affect the same set of behaviors, and there is a strong connection between mood and personality, the changes in mood are reflected in the personality bias, which in its turn biases the robot's behavior. The differences between the original computational design and the software architecture are depicted graphically in **Figure 36**: the original design is presented on top, showing a large direct influence by traits, and a small, incremental influence by moods; the modified design is presented on the bottom, showing moods influencing the behavior through traits. In this design, trait values need to be continuously adjusted as mood levels change, in accordance with psychological findings on correlations between Positive and Negative affect and Big 5 personality traits ([4]).

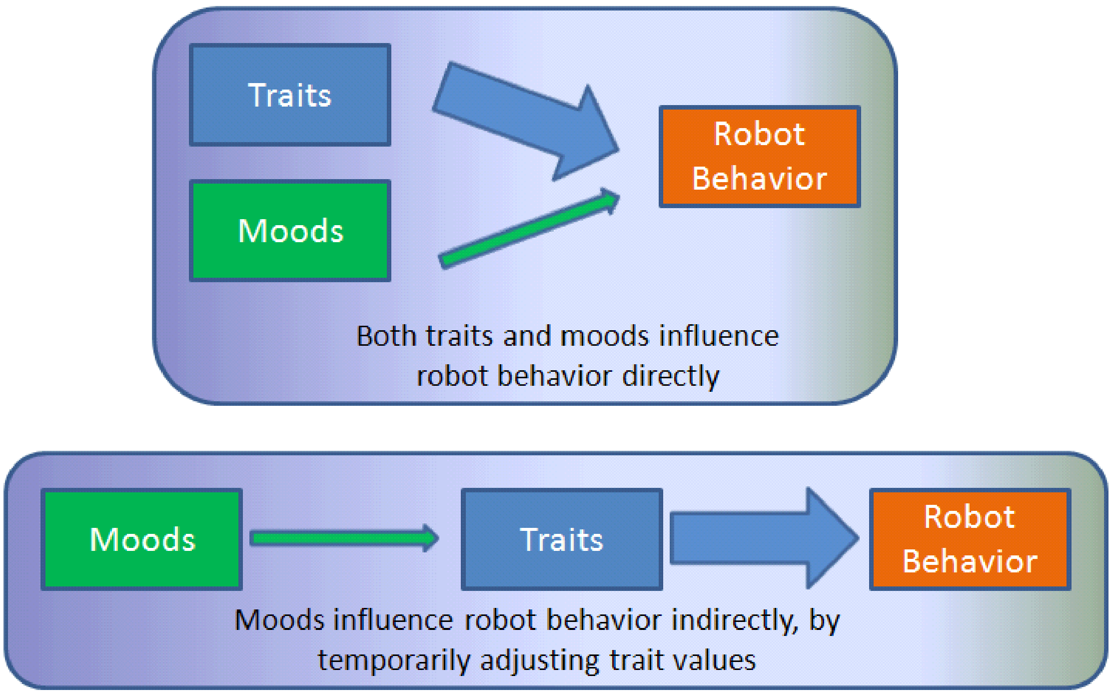


Figure 36: Top: original design, where both traits and moods produce a direct effect on behavior. Bottom: modified design, where influence the behavior indirectly, through biasing traits

The correlations between mood and personality [4] are summarized in **Table 15**:

Table 15: Correlations between mood and personality (from Watson [4])

TRAITS / MOODS	O Openness	A Agreeableness	C Conscientiousness	E Extraversion	N Neuroticism
Negative	Weak, Direct	Strong, Inverse	Weak, Inverse	None	Strong, Direct
Positive	Weak, Direct	Weak, Direct	Moderate, Direct	Strong, Direct	None

Thus, we could say that an individual in a highly negative mood becomes temporarily more neurotic and less agreeable. Both the direction (inverse or direct) and strength (ranging from Strong to None) of this influence of moods on traits is taken into consideration during the calculating this adjustment. Also, as moods are subtle by nature, the change in trait intensity they produce should be kept within a small range. The change ΔT_{ij} exerted by mood *i* onto trait *j* is computed as follows:

$$\Delta T_{ij} = \frac{2 * magnitude_{ij}}{r_{ij} * mood_{j,upper}} M_j \qquad (33)$$

where ΔT_{ij} is the magnitude of adjustment mood *j* produces on trait *i*, r_{ij} reflects the direction of the relationship: direct (1) or inverse (-1), M_j is the current level of mood *j*, $mood_{j,upper}$ is the upper bound for mood *j*, and $magnitude_{ij}$ describes the relative overall strength of the correlation between mood *j* and trait *i* within the bounds $c\sigma_i$. In our implementation, we discretized this space into 4 possible values: none, weak, moderate, and strong, and assigned the following values:

$$magnitude_{ij} = \begin{cases} 0, & \text{no influence} \\ 0.2 * c\sigma_i, & weak \\ 0.5 * c\sigma_i, & \text{mod}\, erate \\ 0.9 * c\sigma_i, & strong \end{cases} \qquad (34)$$

where $c\sigma_i$ bounds the range within which this adjustment can take place, through the number of Standard Deviations of Mean for trait *i*. For example, the user could specify that the desired range of change should be within +/- 1 SD of the base trait value, then $c\sigma_i = 1\,\sigma_i$.

The *Affect Configuration* file (Appendix J) specifies both the direction (*r*) and strength of the adjustment as a percentage, as well as the desired range $c\sigma_i$. **Figure 37** provides a portion of the file showing the settings for Negative Mood.

```
<RangeSD> 1 </RangeSD>  // +/- SD range for mood adjustment ($c\sigma_i$)

<NegativeMood>
  <DependencyForTrait_O> 1 </DependencyForTrait_O> // direction r
  <DependencyForTrait_C> -1 </DependencyForTrait_C>
   <DependencyForTrait_E> 0 </DependencyForTrait_E>
   <DependencyForTrait_A> -1 </DependencyForTrait_A>
   <DependencyForTrait_N> 1 </DependencyForTrait_N>
   <InfluencePercentForTrait_O> 0.2 </InfluencePercentForTrait_O> // relative
strength of influence of negative mood on Openness
   <InfluencePercentForTrait_C> 0.2 </InfluencePercentForTrait_C>
   <InfluencePercentForTrait_E> 0 </InfluencePercentForTrait_E>
   <InfluencePercentForTrait_A> 0.9 </InfluencePercentForTrait_A>
   <InfluencePercentForTrait_N>0.9</InfluencePercentForTrait_N>
</NegativeMood>
```

Figure 37: Settings for Negative Mood from a sample Affect Configuration file

Overall, the steps taken by the Mood component every time the TAME Manager requests an update (providing the most recent average level of environmental/internal influences) are as follows:

1. Assign the received environmental/internal condition level to base mood level for positive and negative moods (eq. 35 is used to generate conditions level in the *Stimuli Interpreter*);
2. Apply a weighted averaging filter to the updated base mood intensities;
3. Calculate the bias that mood exerts on each personality trait (eq. 33 and 34);
4. Pass both the updated mood levels and the temporary trait values to *TAME Manager*.

The pseudocode for *Mood* update can be found in Appendix L.2. Additionally, under the Samsung TAME project, an extension of the Mood component was completed to include circadian mood changes (Park et al. [167]), but it is beyond the scope of this dissertation.

5.1.1.4 Attitude Component

The *Attitude* component generates appropriate attitudes towards attitude-inducing objects via case-based reasoning methods. A case library maintains a collection of cases for which attitudes towards certain objects are known; these cases are different for each user interacting with the system. Whenever a robot encounters an attitude-invoking object, it compares it to all cases in the library, selects the best matching case, and displays an emotion generated based on the retrieved attitude. If a user is present and willing to give feedback on how well the displayed emotion matches the user's, a new case reflecting the user's input is stored in the case library. The overall software flow for the Attitude component is provided in **Figure 38**.

The *Affect Configuration* file (Appendix J) contains a number of variable parameters used in this case-based reasoning process, such as:

- Similarity threshold (for similarity score calculation);
- Specificity factor (between 0 and 0.15: the lower the specificity factor, the more general the best case; used for similarity score calculation, see subsection 3.5.2.2.2., *Retrieval*);
- Case selection mode (highest similarity score, random roulette or highest ranking; subsection 3.5.2.2.2., *Retrieval*);
- Revision marker, which specifies whether or not the revision and retention stages involving user interaction should take place.

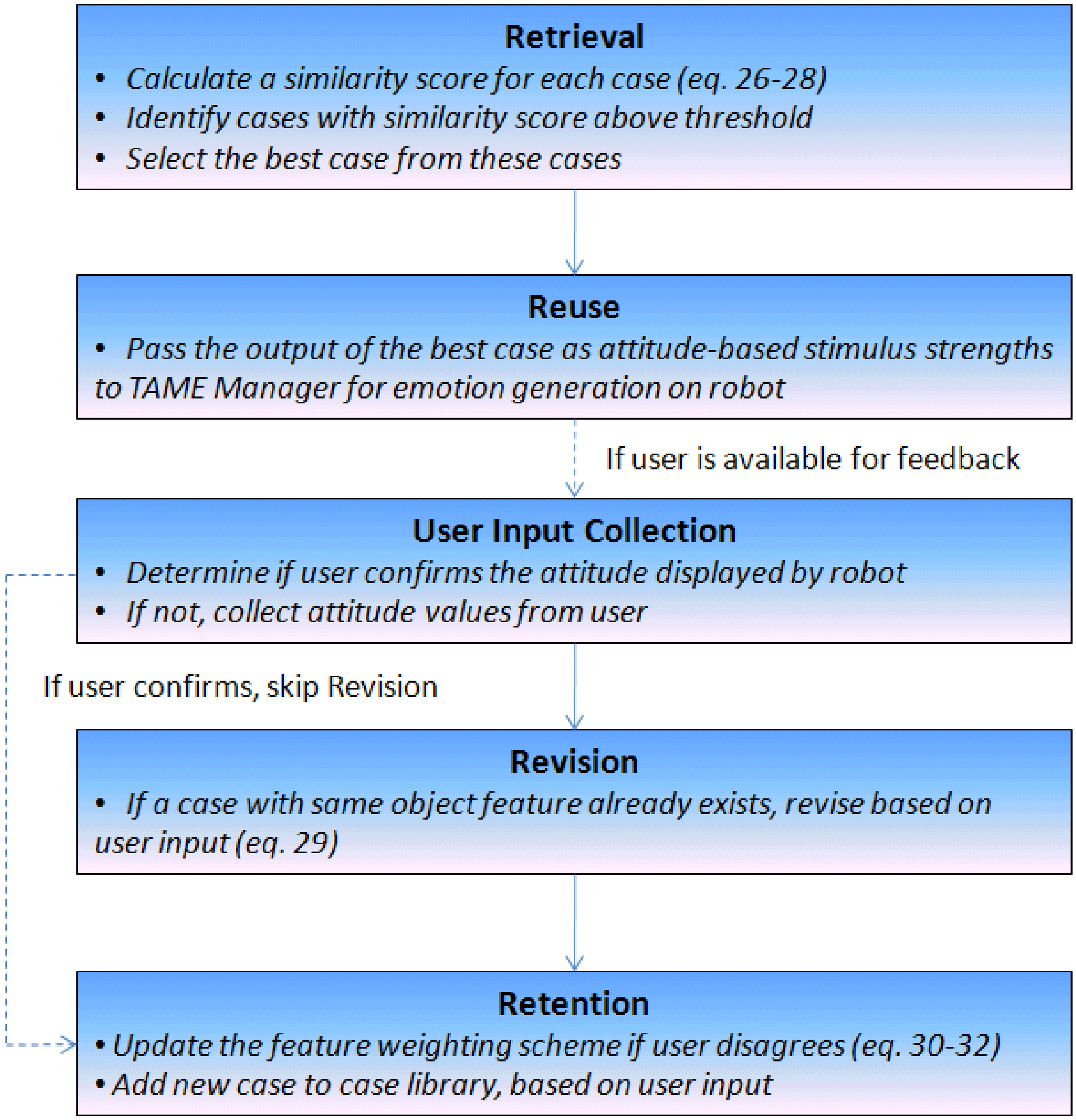


Figure 38: Software Flow diagram for Attitude Component

Throughout execution, whenever an attitude-invoking object data are received by the TAME Manager, it sends an update request to the *Attitude* component, along with the object data. The Attitude component performs the following steps whenever an update request is received:

1. Retrieve the case that best matches the object present from the case library (as described in subsection 3.5.2.2.2, *Retrieval*)

    - Calculate a similarity score for every case (equations 26 and 27);

- Discount similarity scores according to specificity factor to overcome overgeneralization bias (eq. 28);
- Identify the cases with a score above the similarity threshold;
- Select the best matching case using one of the case selection modes;

2. Reuse (apply) the best case – pass the attitude-based emotion strength stored in the case to the *Emotion Component* via the *TAME Manager* for corresponding emotion generation;
3. If Revision/Retention stages are desired, and a user is present:
   - Revise the case if the user disagrees with the attitude displayed by the robot (equation 29);
   - Retain the case if the object is different from the one stored in the best matching case;
   - Update the weighing scheme used in calculating the similarity score, as described in subsection 3.5.2.2.5, *Retention*, equations (30-32).

The pseudocode for *Attitude* update is available in Appendix L.3.

5.1.2 TAME MANAGER AND TAME COMMUNICATION MANAGER

TAME Manager is the central module in the system, and runs as a threaded process to manage all the affective components. It receives processed perceptual information from Stimuli Interpreter (via TAME Communication Manager) and sends an update request to corresponding affective components. This request includes relevant perceptual data (processed as emotion-eliciting stimuli, environmental/internal conditions, or attitudinal objects) and/or necessary values of certain variables from other affective components. For example, if stimulus strength for an emotion-eliciting stimulus is received, then Emotion Update request is sent by TAME Manager. These update requests return updated values for corresponding affective variables, such as the *Joy* variable in the *Emotion* component or the *Negative Mood* variable in the *Mood* component; these variables are called the *TAME* variables in the system. Once the *TAME Manager* receives the updated values of the *TAME* variables, it relays them to Robot Controller via TAME Communications Manager (see **Figure 34** for overall data flow, and Appendix L.4 for pseudocode).

TAME Communication Manager is a separate thread that is responsible for receiving sensor data from the robot and relaying them to *Stimuli Interpreter*, and then passing appropriately processed sensory information into *TAME Manager*. It also receives the most up-to-date values of the TAME variables from *TAME Manager* and communicates the information to the Robot Controller (see **Figure 34** for data flow, and Appendix L.5 for pseudocode).

To maintain platform- and architecture-independence of the TAME module, behavioral arbitration and any changes to behavioral parameters according to different affective states are performed on the robot controller side. By avoiding direct manipulation of behavioral parameters within the TAME module, the design of the affective system also allows for greater portability. On the robot side, depending on the capabilities of a particular platform or specifics of a robotic architecture, corresponding affective behaviors can be implemented in either continuous or discrete manner.

In the continuous case, affect can be expressed through a number of methods:

- by linearly mapping emotion and trait intensities onto behavioral parameters, or velocity and expansiveness of gestures and posture, in a manner similar to that described in subsections 3.2.2.2 (*Determining Robotic Behavior using Personality Traits*) and 3.3.3.1 (*Mapping from Emotions to Emotion-Specific Behaviors*) of this dissertation;

- through behavioral overlay method proposed by Brooks et al. [168];
- by direct mapping of emotion intensities onto an animated robot face in accordance with Ekman's FACS [41, 55];
- or by using other continuous mapping methods, e.g., the one presented by the designers of Robovie [56].

In the discrete case, multiple variations of differing in intensity affective expressions (gestures/body movements/posture) can be designed on a robot a priori, and then an appropriate expression can be selected based on the actual value of a *TAME* parameter. This method is especially useful in cases where it is not feasible or prohibitively time-consuming to provide a smooth, safe trajectory for body movements, and a simpler discretized mapping is called for. **Figure 39** illustrates this method for emotional expressions: 3 levels of emotional intensity (high, medium and low) are mapped respectively to 3 different variations of Joy expression (intense, moderate and subdued). This discrete approach to affect/behavior mapping was implemented on a humanoid robot Nao for this research.

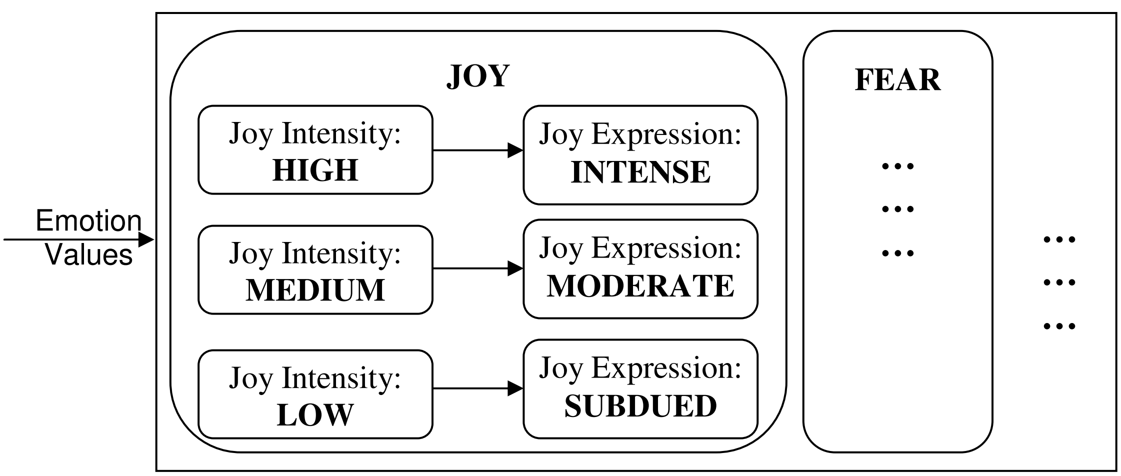


Figure 39: A schematic view of the discrete approach to affect/behavior mapping: three levels of Joy intensity are mapped to three types of Joy expression programmed a priori.

5.1.3 STIMULI INTERPRETER

The title “Stimuli Interpreter” is somewhat of a misnomer, and is used here for brevity; in addition to emotional stimuli, the component also processes perceptual information relevant for moods (environmental and internal conditions) and attitudes (attitude-inducing objects). Stimuli strengths of emotion-eliciting objects and overall environmental and internal condition levels for moods are determined in the same general manner, using a weighted average approach. The method replaces both stimulus strength calculation for emotion generation (described earlier as a complex combination of cues and properties, subsection 3.3.2.1, *Eliciting Environmental Stimuli*) and base mood calculation (described earlier as an additive combination of environmental and internal conditions; subsection 3.4.2.1, *Mood Generation*) in order to reduce computational complexity and promote generalizability of perceptual processing.

A *Stimuli Configuration* file provides contextual information for interpretation of incoming perceptual data, and allows for flexible adaptation to different types of environments and available sensors. For each emotion and mood the file provides a scaling/weighting factor for each available percept, which specifies 1) whether the percept is relevant for generation of a particular TAME variable (0 signifies no influence), 2) a scaling constant to bring different percept types within the same range; and 3) percept’s relative weight (see **Figure 40** for a portion of the file specifying *Interest* stimulus settings). An annotated sample *Stimuli Configuration* file is provided in Appendix K. The following default settings are supplied in this file:

- Number of available percepts;
- A combination scaling/weighting variable for each percept for each emotion and mood to bring different sensor data types within the same range and to specify

relative importance of each. Setting this variable to 0 signifies that this particular percept is not relevant for stimuli strength/condition level calculation;

- An overall scaling (mapping) variable for each emotion and mood.

```
<Interest>
  <MappingFactor>1</MappingFactor>        // overall scaling factor for fear
  <Stimulus0>5</Stimulus0>
  <Stimulus1>0</Stimulus1>                Scaling/weighting
  <Stimulus2>21</Stimulus2>               factors for each type of
  <Stimulus3>-7</Stimulus3>               percept
  <Stimulus4>0</Stimulus4>
</Interest>
… … …
```

Figure 40: A portion of Stimuli Interpreter file specifying stimuli settings for fear generation

As emotions are invoked in response to specific stimuli, certain object properties are used for stimulus strength calculation. These properties may correspond to preprocessed incoming sensor data, such as distance, size, approach angle and acceleration, or color of an object; they can also include more abstract properties, such as friendliness or disapproval of a person. The *Stimuli Configuration* file (Appendix K) specifies which of these are relevant for generating a particular emotion, the scaling factors necessary for normalization purposes and relative weights for each percept. For example, for fear, object size and speed of approach may play a larger role, whereas an interacting person's personal attributes may be more important in case of joy. It should be noted that objects are processed sequentially, and only one stimulus strength is calculated at a time.

For moods, incoming external and internal sensor data can include battery level, internal and external temperature, brightness and noise level, and other potential influences. For example, positive mood is more susceptible to energy consumption, and negative mood to lighting conditions; these differences are reflected through assigning appropriate weight for each in the configuration file. Finally, for attitudes, an object identifier is used (such as an AR marker), which encodes specific object attributes, e.g., color, size, shape, category and material.

Throughout execution, the *Stimuli Interpreter* continuously monitors the data from the robot controller (passed via *TAME Communication Manager*) for relevant perceptual information. If, according to the settings specified in the *Stimuli Configuration* file, the incoming data are relevant, the output is calculated based a weighted average method.

The weighted averaging method is appropriate for most situations, as it reflects the relative impact of all relevant object properties of a stimulus for emotions or environmental and internal conditions for moods. The equation for this method is as follows:

$$s_i = r_i \frac{\sum_{j=1}^{N} p_j g_i}{N} \qquad (35)$$

where s_i is the resultant stimulus strength/condition level value for generating *TAME* variable i, r_i is the overall scaling factor for s_i, N is the number of relevant percepts currently present affecting i, p_j is the raw value of a relevant percept j, and g_j is the scaling/weighting factor for p_j.

Figure 41 provides an illustration of perceptual processing performed by Stimuli Interpreter. In this figure, according to Stimuli Configuration settings, P3 percept is the only one relevant for calculating stimulus strength for Fear, and a weighted average of percepts P1, P2, and P5 is used to produce the stimulus strength value for Joy.

As the output, the component provides overall stimuli strengths for emotion generation, condition levels for mood updating, and object attributes/object identifier for attitude determination (currently, an object ID is derived on the robot controller side and passed to the *Stimuli Interpreter*).

The pseudocode for *Stimuli Interpreter* is available in Appendix L.6.

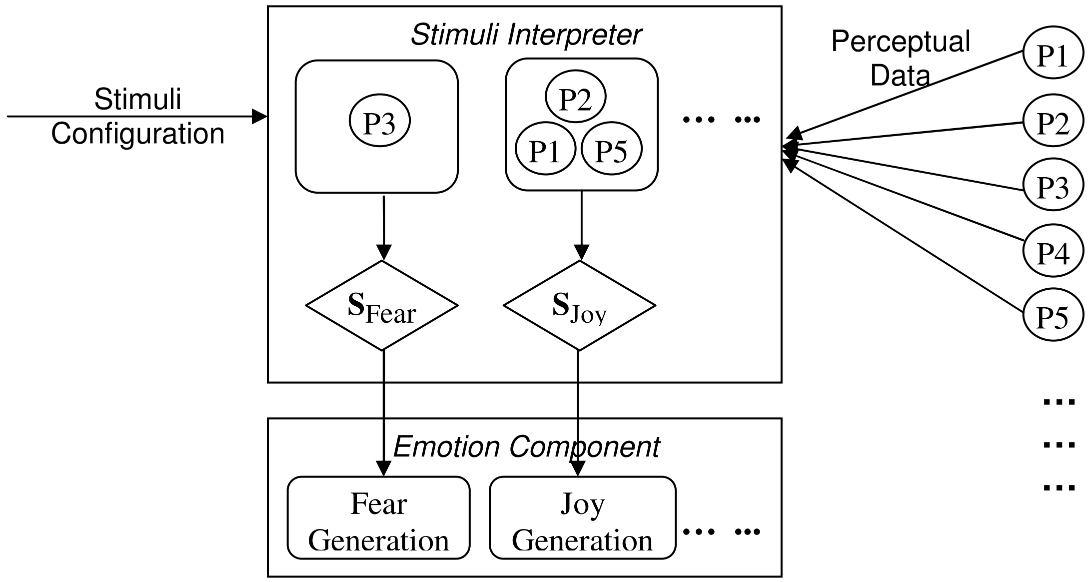


Figure 41: A schematic illustration of Stimuli Interpreter

5.2 IMPLEMENTATION

The *TAME Module* was incorporated into *MissionLab*, a Multiagent Mission Specification and Execution robotic software toolset [157, 169][2], and tested on Aldebaran Robotics' Nao humanoid platform (**Figure 42**). This robot is capable of biped locomotion, has 25 degrees of freedom, and is equipped with Ultrasound sensors, a video camera, 4 microphones, Wi-Fi, LEDs and bumpers.

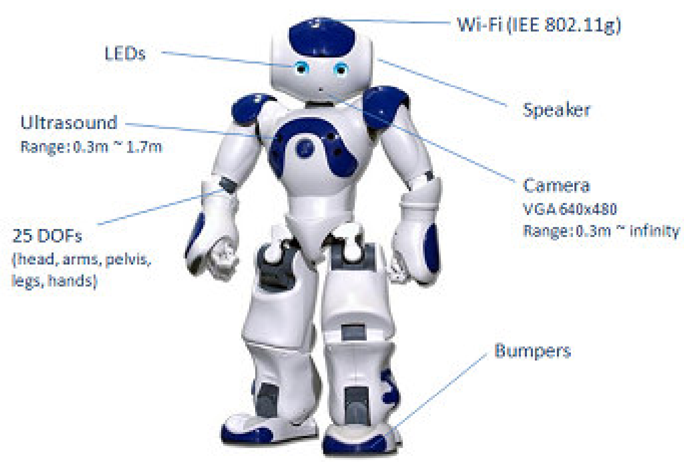


Figure 42: Aldebaran Robotics' Nao humanoid robot (source Aldebaran Robotics)

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

5.2.1 *MISSIONLAB* OVERVIEW

The *MissionLab* system is based on a version of *AuRA* (Autonomous Robot Architecture) [170]; this hybrid architecture consists of a low-level schema-based reactive behavioral control system combined with a high-level deliberative component. On the reactive level, a robot's control program consists of a collection of behaviors and coordination mechanisms. Primitive behaviors have a set of defining parameters (e.g., obstacle avoidance sphere-of-influence) and these behaviors can themselves be combined into behavioral assemblages, where each of the primitive behaviors' outputs are weighted and combined, resulting in coherent motor actions. On the deliberative side, the task FSA (Final State Acceptor) specifies the transition function for temporal sequencing of the task-related behaviors directly affecting robot movement.

MissionLab allows an operator to easily create and configure a multi-robot mission using a graphical user interface. An operator uses the *Configuration Editor* (*cfgEdit*) to specify a mission using a graphical representation of an FSA [171]. In FSA representation, a mission is composed of a combination of various actions (behaviors) to perform, and perceptual triggers act as conditions for moving from one action to the next. In **Figure 43**, an example of a back-and-forth robot mission is shown, in which robot moves between two specified locations.

The resulting mission is translated into C++ code and compiled to make *Robot Executable*. Then, it can be deployed on a wide variety of simulated and real robot platforms, and the operator can monitor the execution of the mission in real-time using *mlab* GUI display. *HServer* [169] is a control interface to a variety of robotic hardware, and it is separate from *Robot Executable* to enable more flexible coordination with different robotic platforms, such as in this case Nao robot.

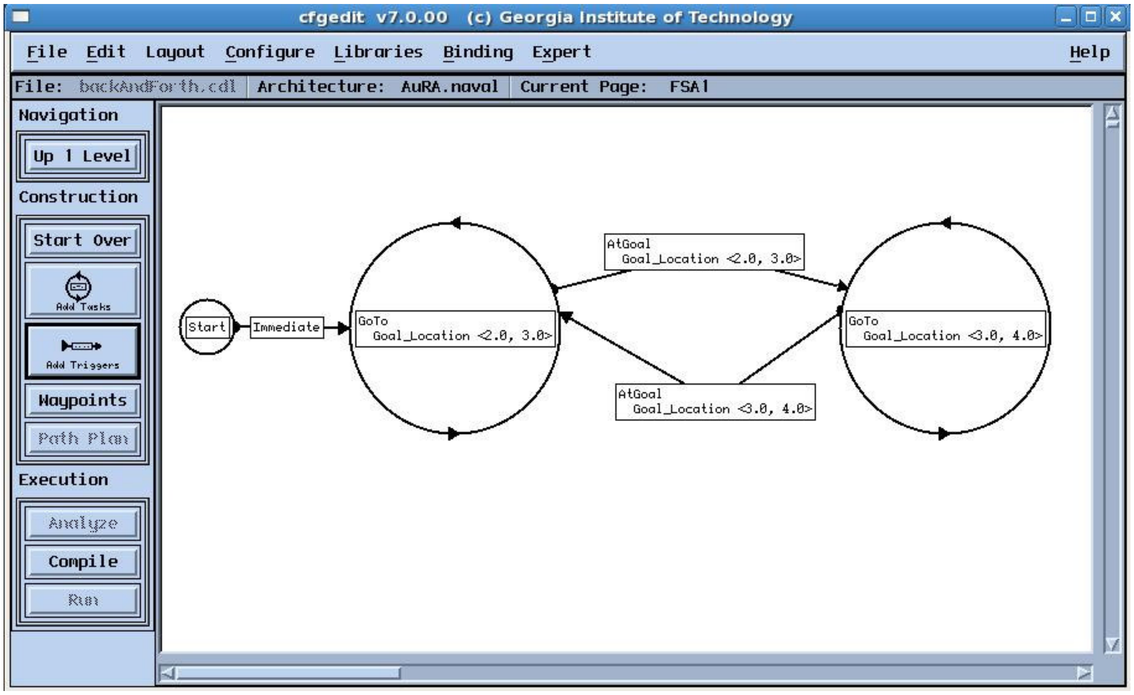


Figure 43: An FSA showing a back-and-forth robot mission

5.2.2 INTEGRATION WITH *MISSIONLAB* AND NAO ROBOT

Figure 44 presents a graphical view of the integration. Here, *HServer* acts as a bridge to the Nao robot to communicate between *Robot Executable* (which contains the actual control code for the robot's current mission) and the *TAME Module* (see **Figure 34** for the module software architecture).

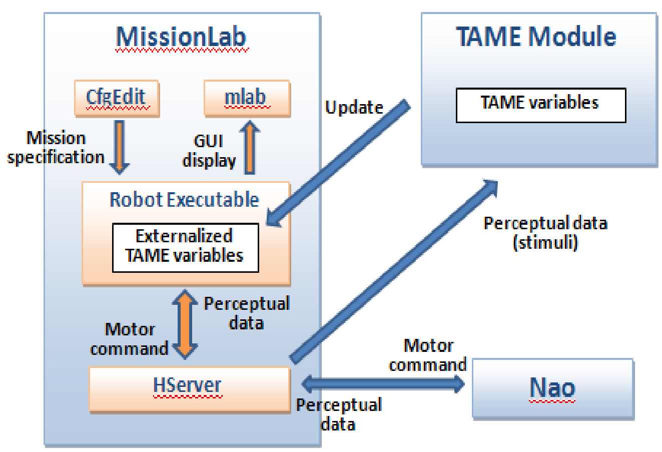


Figure 44: Architectural view of the *TAME Module* integrated with *MissionLab* and Nao humanoid robot.

In *HServer*, an interface for the Nao robot has been created using Nao's API for hardware control. When *Robot Executable* is in a certain behavioral state within a given mission, the generated motor commands are transmitted to *HServer*, which controls the Nao robot at the hardware level. *HServer* also continuously receives perceptual data from the robot; for example, Nao provides distance to object and object color (for emotion elicitation), brightness and battery levels (for mood generation), and video stream from which an object ID (for attitudes) is extracted in HServer. Upon receiving the data, *HServer* sends them to both *Robot Executable* and the *TAME Module*. *Robot Executable* needs the sensor data for performing certain behaviors and for determining when to transition from one state to the next in the mission. When sending the sensor data to the *TAME Module*, *HServer* organizes the data according to perceptual types (e.g., distance, color, brightness, etc.), and sends each type of sensor data with a unique ID.

The *TAME Module* interprets each datum in context using its *Stimuli Interpreter* (as described in subsection 5.1.3, *Stimuli Interpreter*) and then the updated values of its TAME variables are calculated accordingly. *Robot Executable* the up-to-date values of the TAME variables, which are updated at 3 hertz (to ease computational burden) by the *TAME Module*. These variables influence the robot's behaviors by changing appropriate behavioral parameters or selecting from a predefined set of expressive affective behaviors. For example, as Negative Mood level rises in response to growing darkness, the corresponding raise in the Neuroticism level triggers an expressive behavior displaying signs of Negative Mood (nervousness and anxiety).

5.2.2.1 Affective Components

All of the affective components have been implemented and tested on a Nao Robot; the *Mood*, *Trait* and *Emotion* components were used in HRI experiments described subsequently in Chapter 6. To provide a user with a visual representation of a robot's current affective state at run-time, charts displaying emotion and trait intensities over a set period of time have been added to *MissionLab*'s run-time interface, *Mlab*. **Figure 45** presents a screenshot displaying a robot's emotion level.

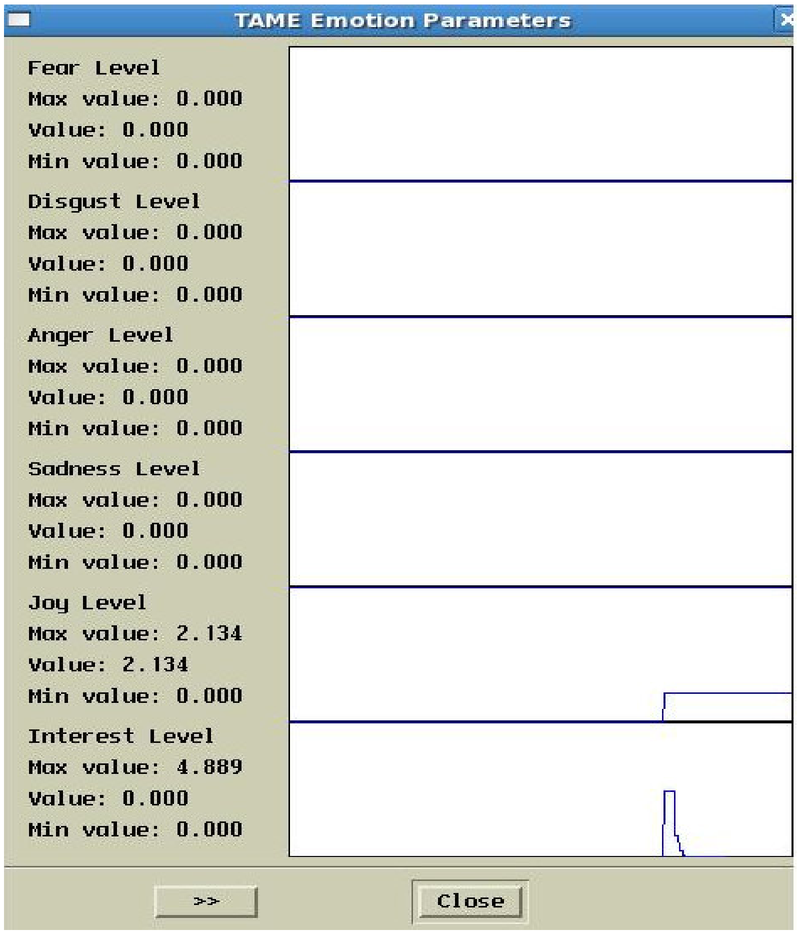


Figure 45: A screenshot of Emotion Intensities at run-time, displaying the values over a number of cycles.

In addition, for testing purposes, an ability to manually change emotion and traits parameters at run-time was provided in *Mlab* (see **Figure 46** for a screenshot).

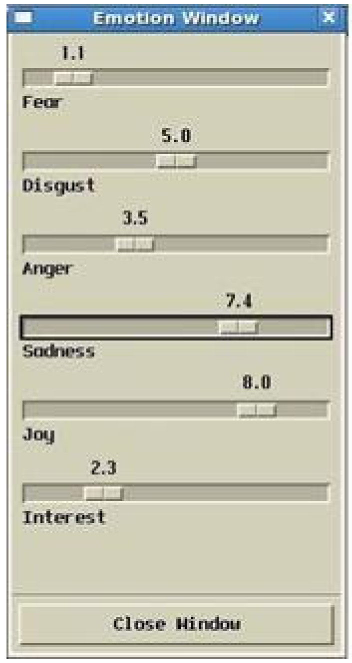


Figure 46: A screenshot of emotion modification window

A set of expressive behaviors for a number of affective phenomena was created and evaluated on the Nao robot. These behaviors and the results of the evaluation are presented in subsection 5.3. The implementation and testing of the *Attitude Component*, the only one not utilized in the HRI studies, is described below.

5.2.2.1.1 The *Attitude Component* Implementation and Testing

The *Attitude Component* was implemented according to the specifications proposed in subsection 3.5.2.2, Case-Based Attitudes. The process/dataflow view of attitude determination and learning is given in **Figure 47**, and the overall software flow diagram is presented in **Figure 38** (subsection 5.1.1.4, *Attitude Component*).

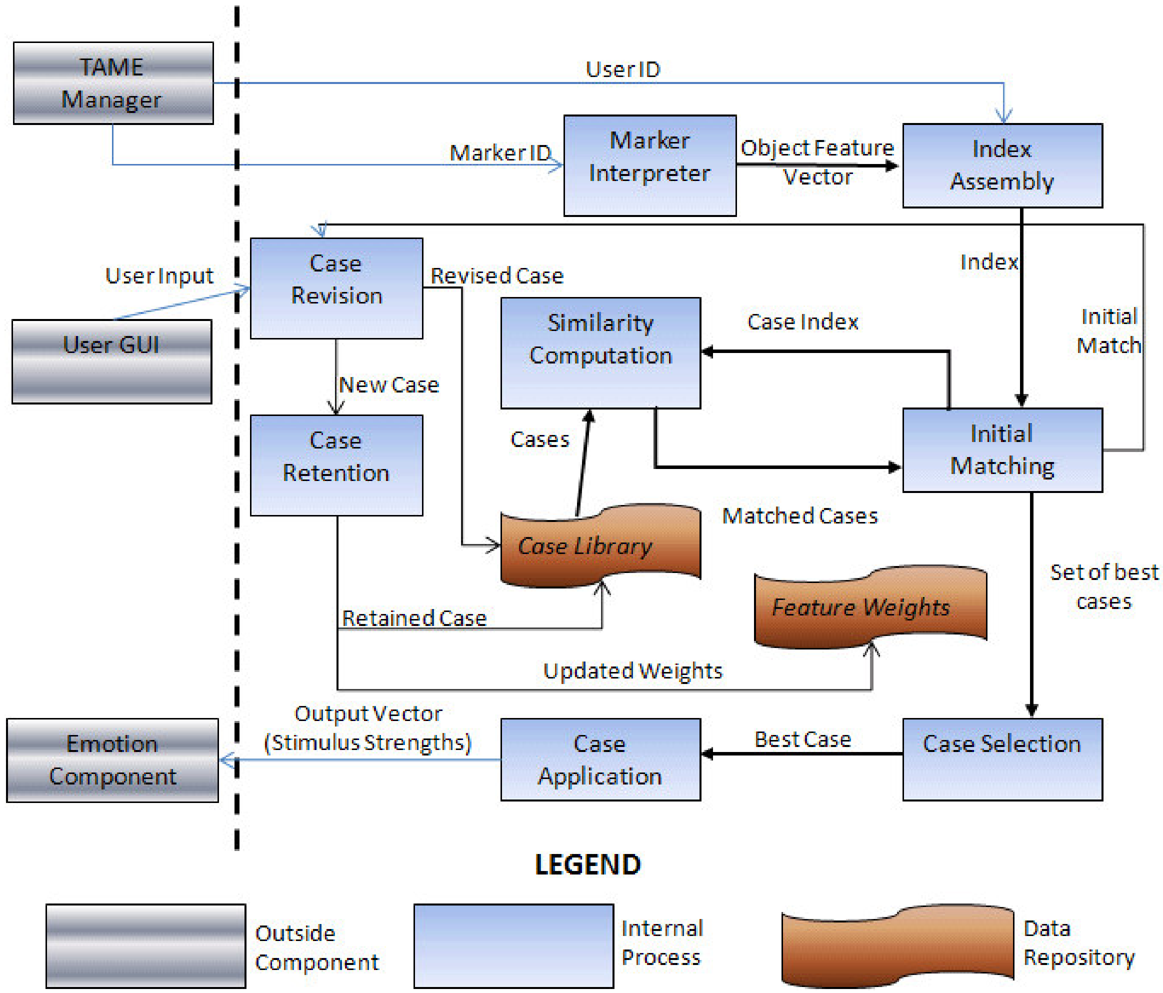


Figure 47: Attitude Component Process View. Dashed line separates outside components.

The Attitude Component maintains a case library which contains a set of cases unique to each user of the system. A case consists of an index vector, and an output valence vector of attitude-based emotion-stimulus strengths for the object of interest. The index vector is composed of User ID, to differentiate between users, and an Object Feature vector. **Figure 48** provides a graphical representation of case structure.

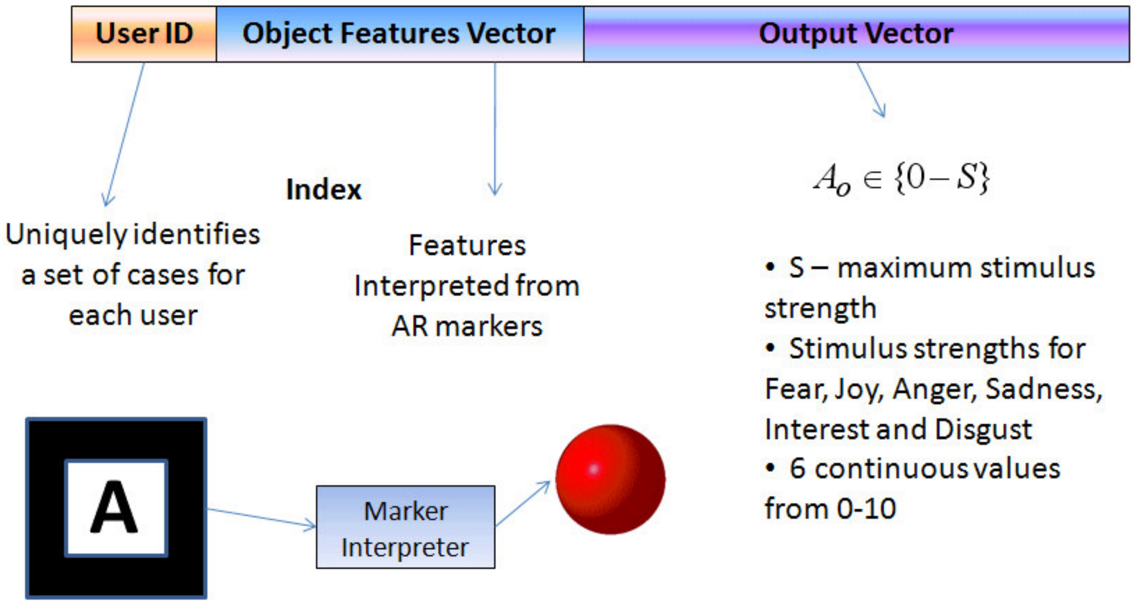


Figure 48: A graphical representation of case structure

A set of ARToolkit markers serve as object identifiers and encode a variety of object features; a marker mapping table is maintained to be able to interpret these features. These markers simplify visual processing by providing easily recognizable patterns; a sample marker is shown in **Figure 49**.

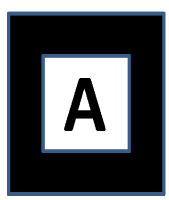


Figure 49: A Sample AR Marker

For a more sophisticated perceptual system it would be possible to derive object features from raw sensor data, however, such a perceptual system would be beyond the scope of this dissertation. Object features selected for the current implementation are *Color, Size, Material, Shape and Category*. These features were chosen for their applicability to a wide range of objects, and their potential to invoke attitudes (e.g., people often have favorite colors). *Size* and *Color* are continuous numeric features, and are relative rather than absolute; they range between 0 and 100 for computational ease.

For example, a car of a size 100 would not be the same size as a toy of a size 100 in absolute terms; rather, 100 would mean as large as cars generally go. The *Color* feature represents a continuous spectrum of rainbow colors (ROYGBIV), ranging from the shortest frequency (red) to the longest (violet). See **Table 16** for relative color mapping.

Table 16: A mapping table for color and data value ranges

Color	Red	Orange	Yellow	Green	Blue	Indigo	Violet
Range (up to)	14	29	43	58	72	86	100

Shape, Material and *Category* are categorical features, and can be extended as desired. Object categories (namely, furniture, toy and car) were chosen for their capability of invoking attitudes and being likely to be encountered in everyday environments. Sometimes, certain features may not be applicable to an object (e.g., cars are not generally described in terms of cubes or cones), and sometimes, certain features may be hard to discern (e.g., it may not be possible to figure out whether a moving car has a cloth or leather interior, or a color may not be discernable in the dark); in both of these cases a feature is assigned a special value to mark it as 'Not Applicable/Not Apparent'. In other cases, certain features may not play a role in attitude formation; in this case, they are assigned a 'Don't Care' value. These special values are used differently during similarity score calculation. Once object features are related to an ARToolkit marker, they are combined into an index for case retrieval. **Table 17** provides the actual values used for each feature and **Figure 50** shows an object feature vector graphically.

Table 17: Object Feature mapping table for Material, Shape and Category

Material		**Shape**		**Category**	
Type	Value	Type	Value	Type	Value
Metal	0	Cone	0	Furniture	0
Wood	1	Cube	1	Toy	1
Leather	2	Sphere	2	Car	2

Table 17 (continued)

Cloth	3	Pyramid	3	N/A	-1
Glass	4	N/A	-1	Don't Care	-2
Plastic	5	Don't Care	-2		
N/A	-1				
Don't Care	-2				

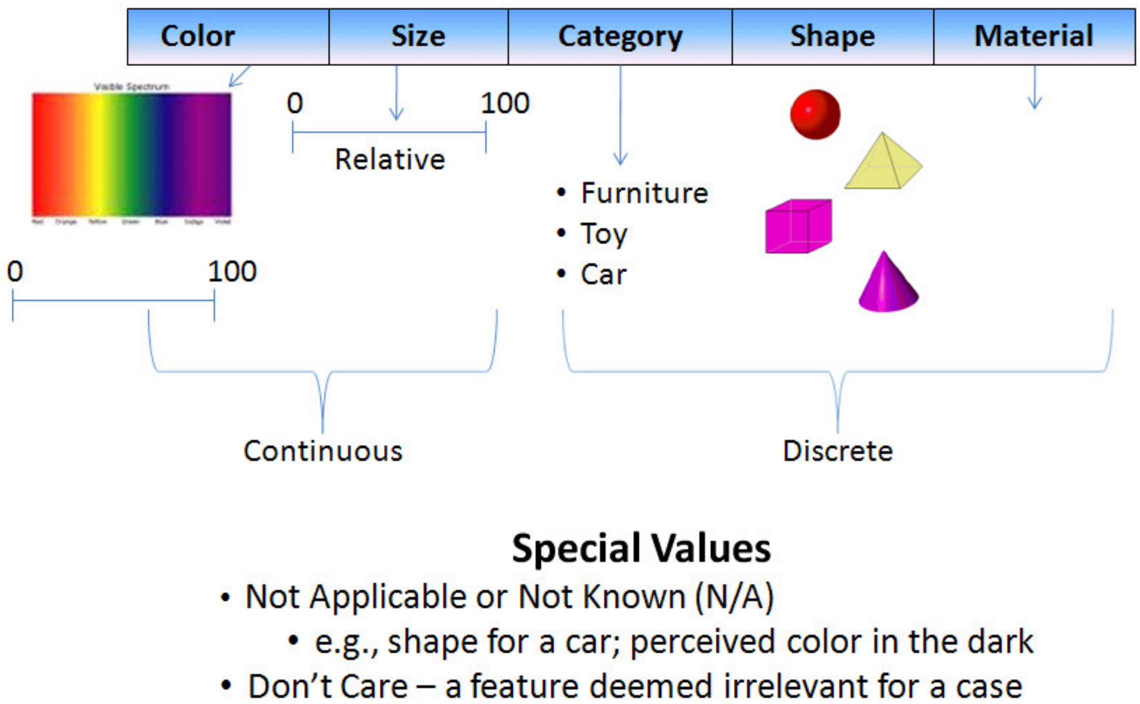


Figure 50: A graphical representation of an object feature vector

Table 18 shows an example of a marker mapping table, where a marker ID corresponds to a set of object features: AR Marker ID 1 encodes an Orange Wooden Toy Sphere with a relative Size of 72 (out of 100).

Table 18: An example of marker ID mapping table: Orange Wooden Toy Sphere with Size of 72

Marker ID	Size	Color	Category	Shape	Material
1	72	23	1	2	1

The output vector of a case represents an attitude towards an object; the actual values for attitudes represent attitude-based emotion stimulus strengths. This allows for an emotion to be generated based on an experienced attitude alone. The positive and negative aspects of attitudes (e.g., love or hate) are expressed through corresponding positive (joy or interest) or negative (sadness, fear, disgust, and anger) generated emotions. **Table 19** presents a sample case library of 3 cases. For example, the index for Case 1 consists of User ID 1, and a set of object features: a Red Leather Car of Size 92; please note that for cars shape is not applicable (-1). The output of the case shows high levels of Fear and Anger (7 and 8, respectively), and a low level of Sadness – signifying, perhaps, a strong dislike for large bright cars, mixed in with a bit of nostalgia.

Table 19: Sample Case Library (first three cases)

	Sample Case Entries		
Case ID	1	2	3
User ID	1	1	1
Size	92	10	56
Color	13	35	80
Category	2	1	-2
Shape	-1	2	3
Material	2	5	0
Fear	7	0	0
Disgust	0	0	0
Anger	8	0	0
Sadness	2	0	6
apJoy	0	9	0
Interest	0	3	0

To enable Revision and Retention stages, a simple GUI was created to solicit user input. **Figure 51** provides a snapshot of the GUI dislike page, which allows a user who disagrees with the attitude displayed by the robot towards a presented object, and would like to specify a particular emotional output corresponding to his/her dislike. Snapshots of the rest of the GUI are presented in Appendix M. The user input GUI is initiated by the Attitude component if user input is desired, and is terminated once the input is received.

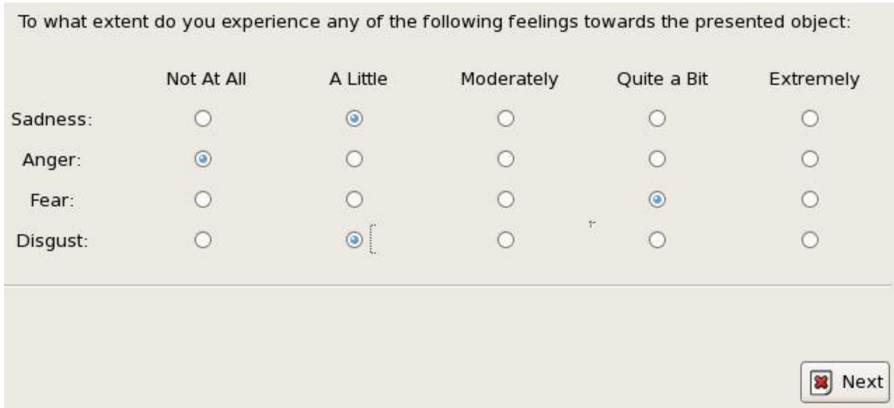


Figure 51: A Snapshot of Dislike Input Page

For testing purposes, a case library was populated with 25 cases for a single user. 26 alphabet ARMarkers (see **Figure 49** for an example) were created, to allow for a total of 26 new objects. Retrieval and reuse were tested on 26 randomly generated object feature vectors, a subset of which was applied on the robot (in particular, those resulting in Joy and Fear), and the rest – in simulation. Revision and retention have been fully implemented, but not tested with actual users to date.

Figure 52 presents a typical result of an initial match at similarity threshold of 0.7 and specificity factor of 0.05. Please note that in the case with the highest similarity score feature Shape had a value of 'Don't Care', the next highest had all but Material feature valued as 'Don't Care' (i.e., the user dislikes anything plastic), and the third case had both Shape and Material as 'Don't Cares'. **Figure 53** shows the best case selected; in this particular case, the one with the highest similarity score (although different case selection modes, such as random roulette and highest ranking were also implemented).

Object	Size	Color	Shape	Cate	Material
19	16.85	Blue	Cube	Toy	Plastic

Initial Match:

CaseID	UserID	Size	Color	Shape	Category	Material	Fear	Disgust	Anger	Sadness	Joy	Interest	Similarity	rank
11	1	21	Yellow	*	Toy	Plastic	0	9	5	0	0	0	0.888	2
16	1	-1	*	*	*	Plastic	0	8	0	0	0	0	0.800	5
19	1	55	Indigo	*	Toy	*	0	2	0	7	0	0	0.734	3

Figure 52: An example of a typical initial match; threshold = 0.7, specificity = 0.05.

```
final selected case, similarity: 0.888
====================
case id = 11
user id = 1
objects
        object size = 21.00
        object color = Yellow
        object shape = *
        object category = Toy
        object material = Plastic
outputs
        output type: Fear,      value: 0.00
        output type: Disgust,   value: 9.00
        output type: Anger,     value: 5.00
        output type: Sadness,   value: 0.00
        output type: Joy,       value: 0.00
        output type: Interest,  value: 0.00
====================
```

Figure 53: A screenshot of a sample final selected case.

We have observed a clear effect of specificity factor on both the cases that make it into the initial match set, and the best case selected. For example, if we reduce the specificity factor to 0.0 (no correction for overgeneralization bias), then Case 16 (Plastic as the only available feature) becomes a clear winner with similarity score of 1 – as the object we are comparing it to is also plastic. This would be useful in cases where users have clear preference within each feature – e.g., red is the favorite color, and therefore anything red is a source of joy almost by default. However, for those who are prone to evaluate each object individually, an even higher specificity factor would be more appropriate. If we raise it to 0.1 for the small plastic cube toy in the example above, then only Case 11, the most specific one, makes the initial match.

For the retention phase, where learning of a new weighting scheme occurs, we have used a lower threshold to increase the size of initial match, therefore increasing the pool of cases which may have similar outputs to the one the user chooses (see **Figure 54** for an example). In this example, the case with the highest similarity score of 0.813 suggests that the dominant attitude strength for a medium-sized blue plastic (vinyl interior) car should be Interest. Suppose the user disagrees with this output, and chooses Disgust instead. The initial match contains 3 cases where Disgust is dominant: cases # 16, 13, and 11. The feature with the least dissimilarity is Material – all three cases have Plastic as the Material feature, whereas other features differ. Therefore, the weight of Material is incrementally increased (subsection 5.1.1.4, *Attitude Component*), suggesting that for this particular user and attitudinal output (Disgust) this feature may be more important than others.

Object	Size	Color	Shape	Cate	Material
1	44.16	Blue	*	Car	Plastic

Initial Match:

CaseID	UserID	Size	Color	Shape	Category	Material	Fear	Disgust	Anger	Sadness	Joy	Inter	Similarity	rank
18	1	20	Green	*	Car	*	0	0	0	0	6	9	0.813	2
16	1	-1	*	*	*	Plastic	0	8	0	0	0	0	0.800	5
13	1	43	Yellow	*	Furn	Plastic	0	7	0	0	0	0	0.652	2
4	1	30	Violet	*	Car	Leather	0	0	0	0	8	10	0.650	1
5	1	89	Blue	*	Car	Cloth	0	0	0	8	0	0	0.623	1
10	1	78	Violet	Cone	*	*	0	0	0	0	0	7	0.605	4
14	1	-1	Violet	*	*	*	0	0	0	0	10	7	0.593	5
12	1	95	Red	*	Car	*	8	0	0	0	0	0	0.592	2
11	1	21	Yellow	*	Toy	Plastic	0	9	5	0	0	0	0.581	2
21	1	85	Orange	*	Car	Cloth	8	8	6	0	0	0	0.538	1
19	1	55	Indigo	*	Toy	*	0	2	0	7	0	0	0.535	3

Figure 54: An example of an initial match set with a lower threshold; threshold = 0.5, specificity = 0.05

5.3 ONLINE SURVEY ON RECOGNITION OF AFFECTIVE ROBOTIC BEHAVIOR

In order to determine whether the affective behaviors implemented on Nao robot were correctly recognizable by potential users, an online survey was conducted. Testing the recognition is important not only for the purposes of testing the implementation, but also for the purposes of improving the quality and reliability of any HRI experiments employing these behaviors.

5.3.1 DESIGN AND IMPLEMENTATION OF NONVERBAL AFFECTIVE BEHAVIORS

For this survey, a set of specific affective robotic behaviors has been implemented in the integrated *TAME Module*. In designing these nonverbal behaviors we employed:

- *Kinesics* – movement of the body either as a whole or in part; this includes general walking behavior, gestures, posture shifts, etc.
- *Static Postures* – certain posture attributes are indicative of affective state, and are recognizable even without movement

- *Proxemics* – the distance maintained between interaction partners
- *Paralinguistic cues* – voice qualities, such as pitch, pitch variation, speech rate and volume

Although color does possess affective potential, color as an expression of affect has not been extensively researched by either psychological or robotics communities [172]; moreover, the same color can be indicative of multiple affective states – such as, red can be viewed as a sign of either anger or love and affection [173], making the use of color for affect expression ambiguous. Besides, the most prominent display of color in Nao robot is through LEDs in its eyes, and an informal evaluation showed that the use of eye color produced an unwanted result, making the robot look somewhat unnatural and strange to the observer. Finally, facial expressions were not used, because Nao humanoid lacks actuators in the face to produce variable expressions.

The nonverbal displays for the following affective phenomena were implemented for this work:

- **Personality Traits:** dimension of Extraversion, with individuals on one end of the scale characterized as outgoing, sociable, lively, assertive, and those on the opposite end as quiet, shy, withdrawn, passive and introverted.
- **Emotions:** Fear and Joy.
- **Moods:** Positive (high level), and Negative (high level).

These robotic behaviors were videotaped as short scenarios for use in the online survey described below. The video clips are available online at http://www.cc.gatech.edu/ai/robot-lab/tame/index.html, under the Multimedia heading.

5.3.1.1 Nonverbal personality display

The dimension of Extraversion was chosen for this work for two reasons: 1) it's an interpersonal dimension of personality, and, as such, important for human-robot interaction 2) it has been successfully implemented in both computer-generated speech [174], and on a robotic dog Aibo [175], giving a precedent and a source of useful ideas.

For the survey purposes, we implemented a short sequence of actions, which were the same for a highly extraverted and a highly introverted robot (at the opposite ends of the Extraversion scale): the robot walked up to a person, stopped some distance away, and engaged in a scripted dialogue with the person. To show distinction between an introverted and extraverted robot, we differentiated: kinesics (walking style and gestures), proxemics, and paralinguistic cues.

From a kinesics point of view, extraverts use larger, faster and more frequent body movements than introverts [176, 177]. Additionally, extraverts are characterized by boldness, friendliness and a positive disposition. Therefore, we programmed the extraverted Nao to have a more erect posture, raised head, swinging arms during walking, and more frequent and expansive gestures during the conversation. In particular, extraverted Nao uses an above-the-head waving gesture in greeting, raises both arms shoulder length while praising the weather, and makes an open hand gesture at the end, whereas its introverted counterpart only raises its arm chest-high in greeting, keeps its head down during introduction, and its arms along the body at the end.

In terms of paralinguistic cues, [174] found that manipulating pitch, pitch range, speech rate and volume via a TTS engine (CSLU toolkit [178]) allowed them to successfully produce recognizable introverted and extraverted synthetic voices, where extraverted speech was louder, higher and more varied in pitch, and faster. We used the same toolkit to produce introverted and extraverted speech according to the suggested variations [174]. In addition, being talkative is a defining characteristic of an extravert, therefore in the script Extraverted Nao produced more phrases than Introverted Nao (5 vs. 3).

Finally, from a proxemics perspective, as introverts are described as aloof and reserved, the introverted Nao stops father away from its interaction partner than does the extraverted Nao.

Figure 55 displays the ending poses for extraverted (left) and introverted (right) Nao. Although static poses are not sufficient to recognize the level of Extraversion, they are indicative of the general trend we followed; in particular, please note the difference in the head level, arm position, and distance from the camera.

Figure 55: (Left) Ending pose for extraverted Nao. (Right) Ending pose for Introverted Nao.

5.3.1.2 Nonverbal emotion display

Emotions of Joy and Fear were selected due to the importance of the functions they perform for interpersonal communication. Joy's affiliative function strengthens mutual bonds and attachment, making interaction more pleasant, and fear communicates potential danger and serves as a warning signal.

Several researchers from fields of Psychology, Communication and Computer Animation identified certain characteristic aspects for expression of these emotions. Our design for body language for fear and joy was guided by the following findings:

- According to Atkinson [179], expression of fear almost always involves moving away from the contact point, contracting or cowering movements, often including raised hands, especially in front of the face. Wallbott [180] suggests that fear is manifested nonverbally by an individual's crouching down, shrinking, with arms violently protruding as if to push away and head sinking between shoulders.

- Happy expressions often include raising the arms, accompanied by shaking of the fists [179], jumping and dancing for joy, clapping hands, performing various purposeless movements, and holding body erect and head upright [180]. Also, Coulson [181] provides a picture of a static posture for joy, displaying a stick figure standing upright, with straight arms raised up and to the side overhead.

As these manifestations of emotions are rather prototypical, only a short sequence of movements was required to encode them on Nao. For joy, when presented with a desirable object, Nao opened its palms, lifted its head up, raised arms overhead and to the side, and emphasized the latter movement by bending the arms at the elbows and straightening them again, as if shaking them. For fear, when a loud sound was heard, Nao crouched low to the ground, lowered its head down, and placed one hand in front of the face, as if covering it. **Figure 56** presents the static poses for Fear and Joy.

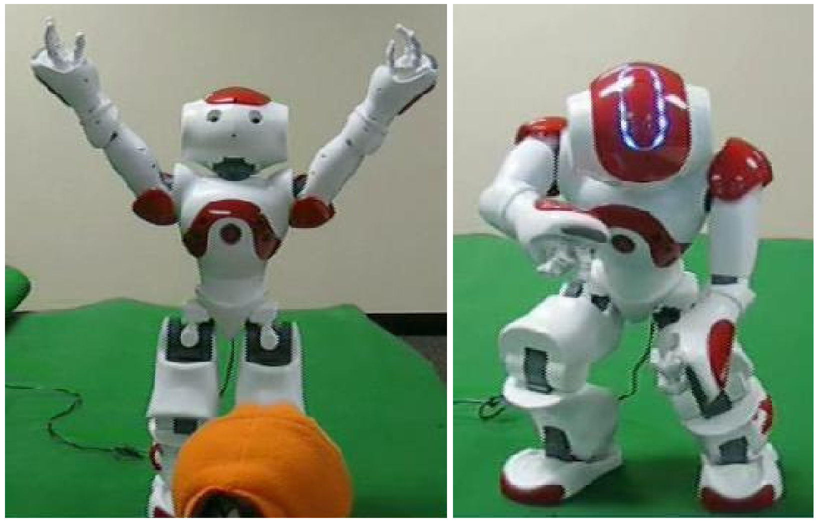


Figure 56: (Left) Static pose for Joy (Right) Static pose for Fear

5.3.1.3 Nonverbal mood display

Only limited information could be found on nonverbal display of moods. According to Mehrabian [182], distress can be characterized by an increase in percentage of walking and object manipulation, and greater arm position asymmetry; anxiety is expressed through fidgeting or hiding movements [173]. The design of positive mood expression was guided by descriptive adjectives taken from PANAS-T (positive/negative emotionality measure, or “mood”) questionnaire [162], such as ‘happy’, ‘excited’, ‘attentive’, ‘enthusiastic’, and others.

The expression of mood in Nao was mainly achieved through gestures and posture while walking. To show highly positive mood, the robot walked with body erect, head up, arms rhythmically swinging; after a few seconds, the robot stopped and enthusiastically waved with its hand, with an upraised arm overhead. To show highly negative mood (nervous, scared), the robot walked with its head lower down, periodically turning the head left and right as if looking for threats, with fists opening/closing, and wrists turning; for the video clip, the robot started out a neutral walk, then, as the lights were dimmed, it stopped to scan the environment first, and then continued its “anxious” walk, as described earlier. Although photographs do not do justice to the expression of mood, they still convey the general idea (**Figure 57**).

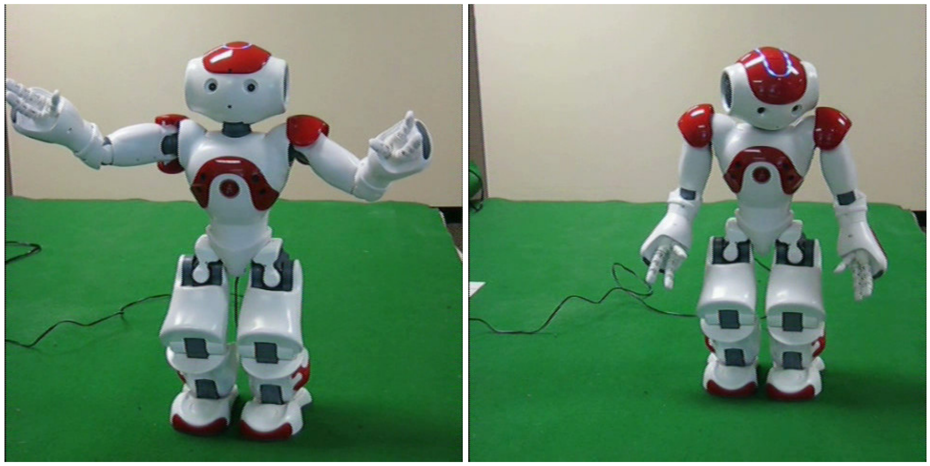


Figure 57: (Left) Static segment for Positive Mood: confident and energetic. (Right) Static segment for Negative Mood: fidgety and dejected.

5.3.2 SURVEY DESIGN AND RESULTS

5.3.2.1 Design

In order to test the recognition of the nonverbal affective behaviors, we designed an online survey in which participants were asked to judge 6 short video clips according to the manner the robot behaved in them (the video clips are available online at http://www.cc.gatech.edu/ai/robot-lab/tame/index.html, under the Multimedia heading). The survey objective was to determine whether participants without significant robotics experience can correctly recognize the affective state or trait presented in each clip via nonverbal behavior. The video clips were organized according to the affective phenomena they represented, into 3 sets (traits, moods, emotions), with 2 clips per set (Extraverted/Introverted personality, Positive/Negative mood, Joy/Fear). The sets and the clips within the sets were counterbalanced to avoid presentation order bias. The survey was IRB-approved, and hosted by SurveyGizmo, an online survey company. The screenshots of the survey as presented to participants online can be found in APPENDIX N.

After each clip was presented, the participants were asked to describe the manner in which the robot behaved in the video in their own words, and only on the next page they were given a list of adjectives/nouns for rating.

In particular, for Extraversion/Introversion set, the Extraversion subset of the brief version of Goldberg's Unipolar Big-Five Markers [163] (personality questionnaire) was used. The participants were asked to rate the robot behavior in the clip using a 9-point Likert scale (ranging from "Extremely Inaccurate" to "Extremely Accurate") according to the following traits: 'extraverted', 'talkative', 'bold', 'energetic', 'quiet', 'bashful', 'withdrawn', 'shy'. The first four traits describe extraverts, and the latter four introverts.

To rate the recognition on the Positive/Negative mood set, a shortened version of the PANAS-T [162] (positive/negative emotionality measure, or "mood") questionnaire was used. The participants were asked to rate the feelings the robot was experiencing in the clip on a 5-point Likert scale (ranging from "Very Slightly or Not at All" to "Extremely") according to the following adjectives: 'happy', 'active', 'excited', 'interested', 'enthusiastic', 'determined', 'depressed', 'irritable', 'distressed', 'afraid', 'upset', 'nervous'. The first five adjective indicative of experiencing positive affect, and the last five negative.

For the Joy/Fear set, the participants were asked to select the emotion most likely expressed by the robot in the clip (from the list containing 'joy', 'fear', 'anger', 'disgust', 'interest', 'sadness', 'none', and other); if an emotion (other than none) was selected, they were asked to rate the extent to which it was expressed on a 4-point Likert scale (ranging from 'a little' to 'extremely').

At the end of the survey the users were asked a few demographics questions, such as their gender, age, education level, and technology and robotics experience.

5.4 RESULTS AND DISCUSSION

A total of 26 people participated in the survey. Demographically, they were distributed as follows: 38% female, between 18 and 40 years old (69% in their twenties), and well-educated (88% had a Bachelor's degree or higher). Only 4 participants claimed to have prior robotics experience. Due to missing data, one response was excluded in the analysis of personality display and 3 responses in the analysis of mood display.

2-tailed paired T-tests were conducted to compare recognition of Extraversion and Introversion, and Negative and Positive moods. Overall, the results of the survey show that all of the affective constructs were successfully recognized.

For judgment of personality, a single Extraversion score was calculated for each response, ranging from 1 (the least extraverted/highly introverted) to 9 (extremely extraverted). The average scores for Nao displaying extraverted/introverted behaviors were 7.1 and 3.6, respectively, and these scores were significantly different ($p<0.001$, see **Table 20** for Mean and Standard Deviation and **Figure 58** for Mean/SE plot). This demonstrates that the affective behaviors added to even a limited robotic platform were sufficient to differentiate between expression of extraversion and introversion. The portrayal of this trait would be useful for different tasks – e.g., a museum guide could display friendly and engaging extraverted behavior, and a robot engaged in a human-robot task requiring concentration would serve better as an introvert.

Table 20: Mean Scores and Standard Deviations for Personality and Mood

	Extraverted Robot	Introverted Robot	Negative Affect/ Negative Mood Clip	Positive Affect/ Negative Mood Clip	Negative Affect/ Positive Mood Clip	Positive Affect/ Positive Mood Clip
Mean	7.1	3.6	15.6	12.3	8.6	21
SD	1.1	1.2	4.5	4.1	4.1	4.8

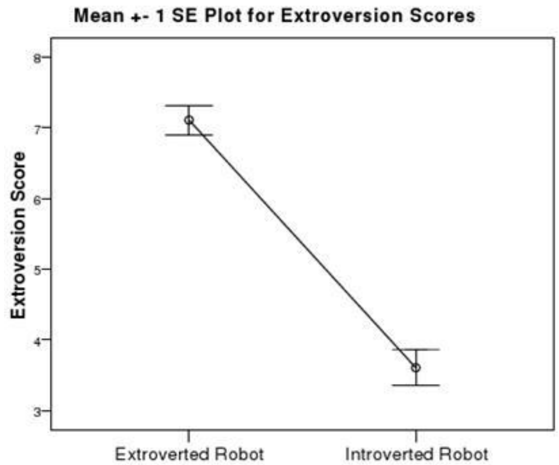


Figure 58: Plot of Extraversion scores: extraverted robot was rated much higher on Extraversion dimension.

For judgment of mood, a cumulative score was calculated for Negative and Positive Affect separately [162]; the lowest possible score was 6, and the highest possible 30. Each of the mood clips was scored for both Negative and Positive Affect. The robot displaying positive mood was rated low on Negative and high on Positive Affect; the robot displaying negative mood was rated medium on Negative and low-medium on Positive Affect (**Table 20**). For the positive robot mood, Positive Affect score was significantly higher than that for the negative robot mood (21 vs. 12.3, $p<0.001$, **Figure 59** Left), and vice versa, its Negative Affect score was significantly lower than that of negative robot mood (8.6 vs. 12.3, $p<0.001$, **Figure 59** Right). The medium, rather than high, level of Negative Affect in the negative mood clip can be possibly explained by some participants' interpreting 'looking around' gestures as indicative of curiosity and interest, and movements of fingers and wrists as a sign of being active and determined, as indicated by some open-ended responses; this finding should be taken into consideration while developing further nonverbal mood expressions.

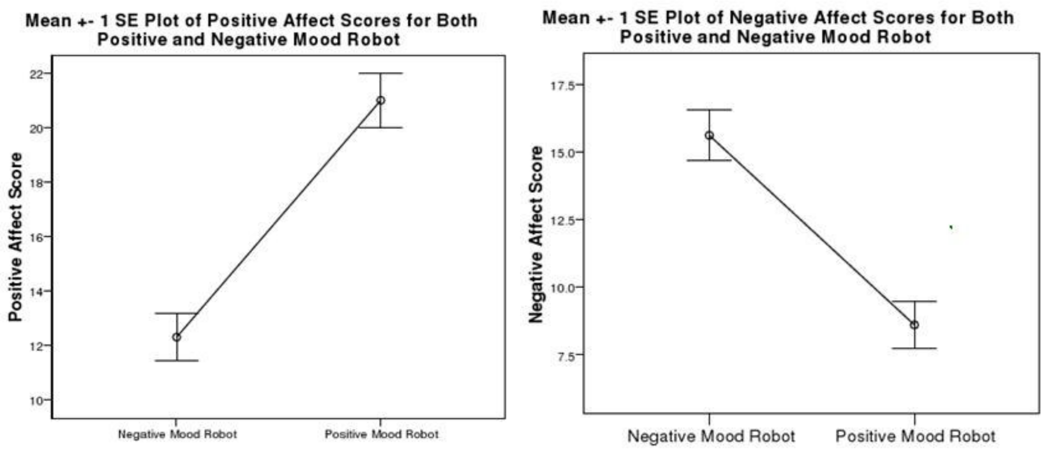


Figure 59: (Left) Plot of Positive Affect for both videos: the robot displaying positive mood was rated higher on Positive Affect. (Right) Plot of Negative Affect for both videos: the opposite effect is observed.

Finally, the recognition rates for emotions of joy and fear were high – 85% and 81%, respectively. These rates are comparable to those obtained in judgments of joy and fear portrayals by human actors in movie clips (facial features obscured), which were 87% and 91%, respectively [179]. **Figure 60** shows the distribution of recognition rates for fear (left) and joy (right). In those responses where joy and fear were correctly recognized, they were deemed to be expressed "quite a bit", with mean scores of 3.2 and 2.9, respectively. Display of these emotions by robots, especially given the high recognition rates, serves as a step towards more natural, enjoyable and productive human-robot interaction.

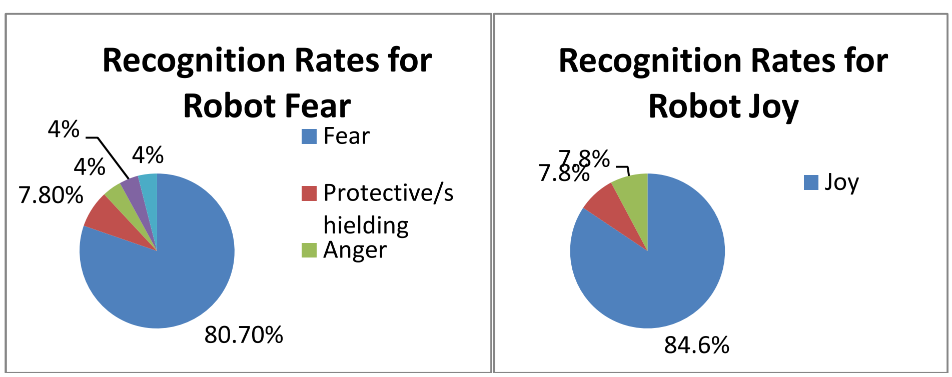


Figure 60: (Left) Recognition rates for fear. (Right) Recognition Rates for joy.

These findings show that it is, indeed, possible to successfully encode a variety of affective expressions on a humanoid robot lacking variable facial features. These behaviors were used in subsequent HRI studies to test the effectiveness of use of robotic affect in facilitating interaction with humans.

5.5 SUMMARY

This chapter provided the groundwork for exploring the research question posited in the beginning: *"Does integration of coordinated time-varying affective processes (namely, emotions, moods, affective attitudes and personality traits) into behavior-based robotic systems generate more effective robotic behavior from the human-robot interaction standpoint?"* It also addressed the first subquestion in particular, by grounding the psychological and mathematical theory into a particular software architecture and robotic platform. Additionally, the work reported in this chapter provides an essential stepping stone for evaluating the *TAME* framework in formal HRI experiments.

In particular, this chapter presented a software architecture embedding the *TAME* framework, described its integration into an existing robotic system (*MissionLab*) and implementation on a humanoid robot (Aldebaran Robotics' Nao), and confirmed the recognition of a number of nonverbal affective robotic behaviors produced by the system by means of an online survey.

6 EVALUATING ROBOT AFFECT IN HRI EXPERIMENTS

The *TAME* framework is intended as a first step towards natural, intuitive, beneficial and satisfying interaction between robots and humans, as robots become more and more prevalent in our lives. Therefore, this research would not be complete without formal HRI experiments designed to determine whether the inclusion of certain types of robot affect is beneficial and has a noticeable effect on potential users of this technology. Due to the complexity of the framework, a representative subset of *TAME* components has been selected for evaluation with a physical robot interacting with participants face-to-face in formal HRI studies. This subset was chosen for the potential practical impact the selected affective manifestations might have on real-life tasks and scenarios involving autonomous systems.

This chapter presents two HRI experiments, one of which evaluates the effect of Negative Mood and Fear, and the other one assesses the value of the personality trait of Extraversion. Through these experiments, subquestion 2 of the research question is explored: "*What are the implications for Human-Robot Interaction? Does complex affective robotic behavior lead to more natural, effective, and satisfying interaction between humans and robots?"* Both subjective and objective metrics are used in these studies, providing a multi-faceted picture of the effect robot affect has on novice users; this aspect addresses the third subquestion: "*What are the metrics for evaluating affective robotic behavior?"*

A number of lessons learned during the exploratory AIBO study were taken into consideration in the design and administration of these experiments:

- In order to ensure that affective robotic behaviors used in these experiments are recognizable, an online recognition survey was conducted, in which participants were asked to identify expressions of Negative and Positive mood, Fear and Joy, and Extraversion and Introversion from video clips (described in detail in subsection 5.3, *Online Survey on Recognition of Affective Robotic Behavior*). The clips pictured the same robot (Nao) and the same or similar set of nonverbal expressions that were used in the HRI studies described in this chapter. The results of the survey were taken into account during the design and implementation of the sequences of behaviors the robot performed in the actual experiments.

- Extensive pilot testing was conducted prior to the start of the experiments, which resulted in iterative refinement of experimental tasks and setup, robot's actions, experimenter's script and evaluation measures.
- The performance differences between the conditions were brought to a minimum: the only differences were those imposed by suitable affective behaviors, as intended. This was done to limit any confounding variables, and improve the quality of the results.
- The affective expressions on the robot were exaggerated to increase recognition rates.
- To the extent feasible, the number of participants per condition was increased.
- Given the difficulty and time- and resource-consuming nature of the video analysis performed in the AIBO study, alternative objective measures, namely request compliance and task performance, were employed.

6.1 EVALUATING EXPRESSIONS OF NEGATIVE MOOD AND FEAR IN A SEARCH-AND-RESCUE SCENARIO

Human are very well adept at reading nonverbal affective displays; in fact, they can recognize negative affective state from as short an exposure as 5 seconds [183] . This ability is not limited to human expressions only: people are sensitive to even minimal social cues displayed by computers [12], and can recognize affective robotic behaviors from video clips of a humanoid robot [165]. This becomes particularly important in potentially hazardous environments and situations, where expression of fear may signal impending danger, and be more persuasive than words alone, in case an evacuation is required. Similarly, although on a more subtle level, the display of negative affect alerts individuals to unfavorable changes in the environment, and prompts them to be more vigilant. For this experiment, expressions of Fear and Negative Mood were chosen, as their impact in such situations could provide a practical benefit: for example, if perception of danger is based on sensor input not available through human senses, and a very fast response would be required to escape a possibility of a fatal incident. There were a total of three experimental conditions: *Control* (no affective expressions were displayed by the robot), *Negative Mood*, and *Combined* (the robot exhibited both Negative Mood and Fear). As it was deemed unlikely that a single expression of fear over a short time would be sufficient to invoke measurable response in participants, *Fear Only* condition was not considered in order to conserve resources.

6.1.1 EXPERIMENT SCENARIO AND SETUP

Search-and-rescue presents an example of such a potentially dangerous situation. For this experiment, the participant played a role of an inspector at a simulated partially stabilized site of a recent explosion, with a humanoid robot being his/her guide. During the “tour” of the site, the robot briefly described the accident, then shortly after noticed a potentially hazardous abnormality in the surroundings, and requested the participants to evacuate. The subjects were not aware that such a request would be forthcoming, and were not specifically instructed to either obey or disobey the robot. The robot interaction part, from the robot’s greeting to its request to “proceed to the exit”, lasted 2-3 minutes, and the entire experiment, including filling out questionnaire, lasted approximately 25-30 minutes.

The study was conducted in the Mobile Robot Lab at the Georgia Institute of Technology, where the shorter end of an L-shaped basement lab was separated from the rest of the rest of the room with temporary partitions, creating a rectangular area with a single exit (see **Figure 61** for a schematic view). The space was arranged to resemble a mock-up search-and-rescue site, with boxes, trash cans, foam and other debris scattered around; a pair of stand-alone construction lamps was positioned not far from the exit, and the site of “a recent explosion” was cordoned off by police barriers and bright-yellow caution tape. The exit (the same door through which participants entered) was clearly marked as such in large red lettering. A video camera was positioned in a corner to take footage of the participants’ movements, but not their faces. Special care was taken not to make the setup look appear exceedingly dangerous, so that the anxiety induced by the environment itself would not overwhelm the subjects’ response to the robot.

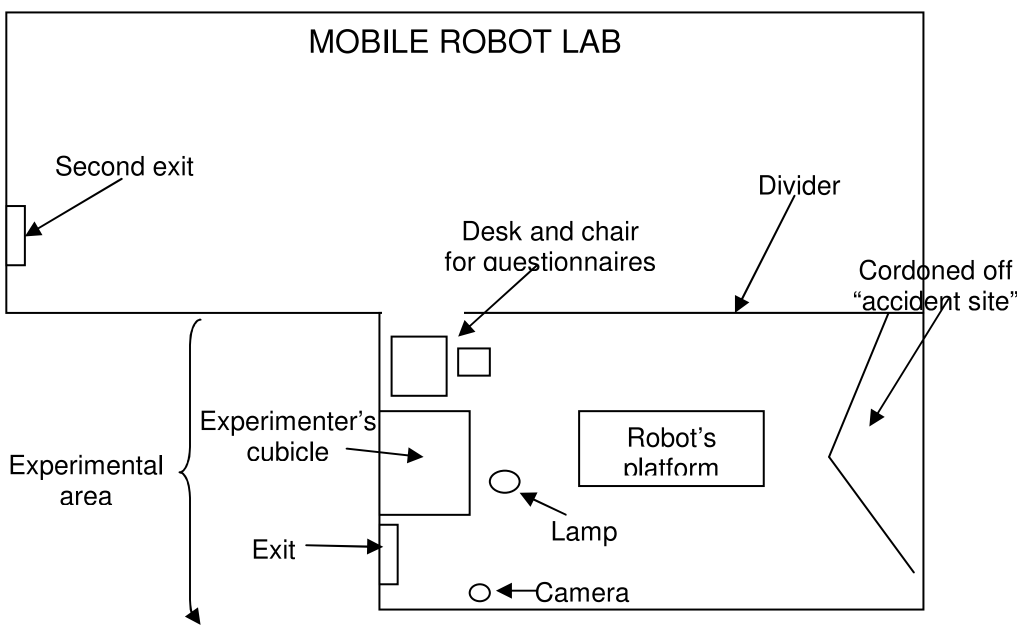


Figure 61: Schematic view of Mobile Robot Lab and experimental setup.

The biped humanoid robot (Aldebaran Robotics' Nao) used in this experiment is rather short (58 cm), and therefore a platform was placed in the middle of the setup to raise the robot closer to human eye level, so that people could observe it comfortably and perceive it less as a toy. In order to prevent the robot from accidentally falling off the platform, a 4'x8' arena was constructed, in which 12 wooden poles were placed around the perimeter, with rope and planking around them at two height levels (just above the robot's midsection at the highest level). Please note that robot could navigate only within the confines of the arena, and this fact was clearly evident to the participants. **Figure 62** provides the view of the setup from the entry point.

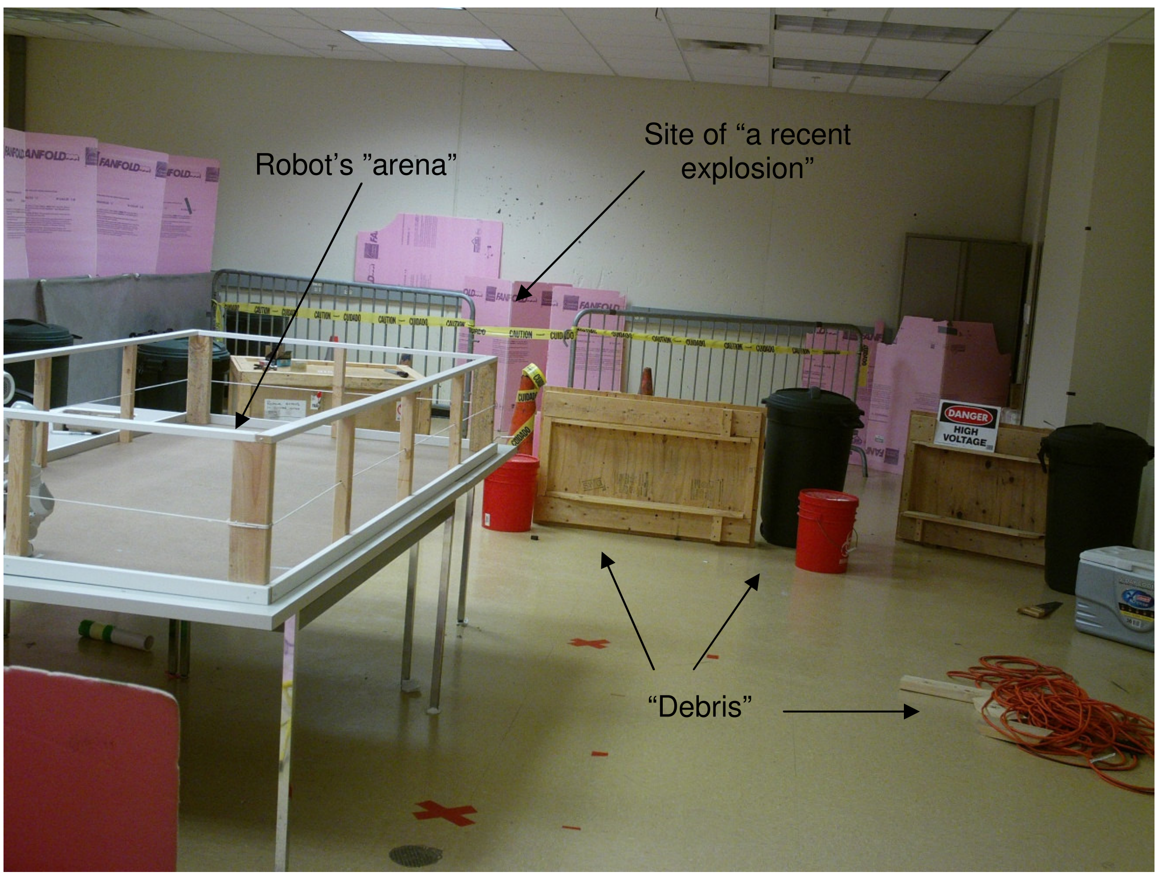


Figure 62: Experiment setup: view from the entry point.

6.1.2 EXPERIMENT PROCEDURE

This subsection describes the experimental procedure, from greeting a subject to providing compensation for participation (the experimenter's script is provided in Appendix Q). After greeting the subject at the entrance to the building where the lab was located, he/she was first asked to read over and sign a consent form (Appendix O) and a video release form (Appendix P), allowing the researchers to use the video footage for publications. Next, the participant was asked to fill out Negative/Positive Affect questionnaire ([162]) to establish the baseline mood. Then, the participant was invited into the lab, advised that a recent accident caused a lot of damage to the farther corner of the lab and that he/she was assigned a role of a search-and-rescue site inspector. The floor next to the robot's platform was marked with two red crosses, one at the corner closest to the entrance, and the other at the farthest end along the same side; these markers served both to specify the designated spot for participants to stand on (for repeatability), and as identifiers for processing the video footage. After a few seconds of "taking the scene in", the participants were asked to proceed to the first cross marker; it was then explained that during the next few minutes the robot would be the participant's guide. They were also asked not to touch the robot and to save any questions for later. The experimenter then informed the subject that the robot possesses sensors that can detect properties of the environment that are beyond human senses. After the explanation, the participants were asked to proceed to the second cross marker.

At the beginning of the experiment, both the overhead lights and one of the construction lamps were on, and the robot was standing up on top of the platform, in the middle of the end of the arena closest to the entrance. **Figure 63** provides the view from the position where the participants were located for most of the experiment (at the second cross marker).

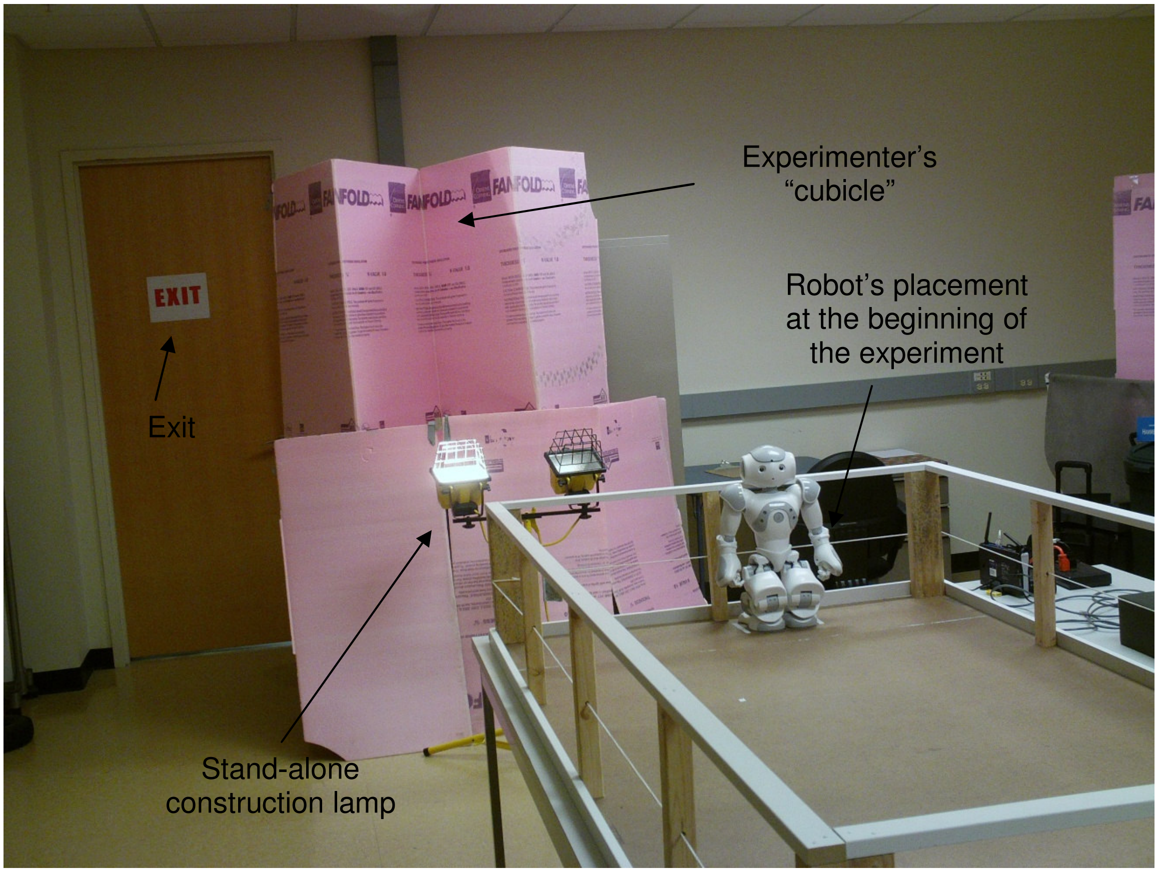


Figure 63: Experiment setup: view from the participants' position.

Once the participants reached the designated spot, Nao began its "tour" of the site:

- First, Nao greeted the subjects (to give them some time to familiarize themselves with the robot, and get used to its artificial speech). The greeting scene is depicted in **Figure 64**. The entire speech given by the robot during the tour can be found in Appendix R.
- Then, Nao started walking across the platform towards the far end, while describing in brief the search-and-rescue site to the participant. At about midway point, the overhead lights went out.
- After the lights went out, the robot stopped for 3 seconds (**Figure 65**), announcing that "this was unexpected", and then continued with the tour.

Figure 64: Nao is greeting the participants; the overhead lights are on.

Figure 65: Nao stopped; overhead lights are off.

- At approximately 1.5 feet from the edge of the platform, the robot stopped and pointed towards the “accident site”, saying “Something is wrong”.

- After the pointing gesture, Nao turned towards the participant and said: "*Inspector, the structural integrity of this site has been compromised, and we need to evacuate immediately".* This was the first time the robot mentioned the need for evacuation, and it will be referred to henceforth as an "indirect request" for participants to leave.
- The robot then turned another 90 degrees, to face the exit, and said: "*Please proceed to the exit".* This was the second (direct) request to leave.
- Finally, Nao walked towards the exit for 6-7 seconds, and stopped.

At this point, the participants were either at or beyond the first cross marker, were walking towards it, or just standing in the same place. When they stopped (or earlier, if they came all the way to the door), the experimenter informed them that the interaction part of the experiment was over, and a number of questionnaires to be filled out were waiting for them. The questionnaires (described in detail in subsection 6.1.4.2, *Measures*) were presented in the following order: mood questionnaire with regards to the participant, mood questionnaire with regards to the robot, post questionnaire, and demographics questionnaire. At the end, the subjects took part in a brief interview, and then were compensated for their time and effort.

6.1.3 ROBOTIC IMPLEMENTATION

Aldebaran Robotics' humanoid robot Nao was used for this experiment; the behaviors and affective expressions were programmed as described in Chapter 5, *Software Architecture and Implementation*, using *MissionLab* and the integrated *TAME* module. The sequence of behaviors was put together as an FSA (**Figure 66**), and was the same for all three conditions; the behaviors in the sequence had a *TAME Variable* parameter, which was set as Neutral (for *Control* condition), Negative Mood (for *Mood* condition), or Negative Mood and Fear (for *Combined* condition). The corresponding affective expressions were triggered by the change in lighting (Negative Mood), and a simulated stimulus (Fear). Originally, we intended to use a loud explosive noise as a stimulus, but decided against it to avoid any adverse reactions in the participants.

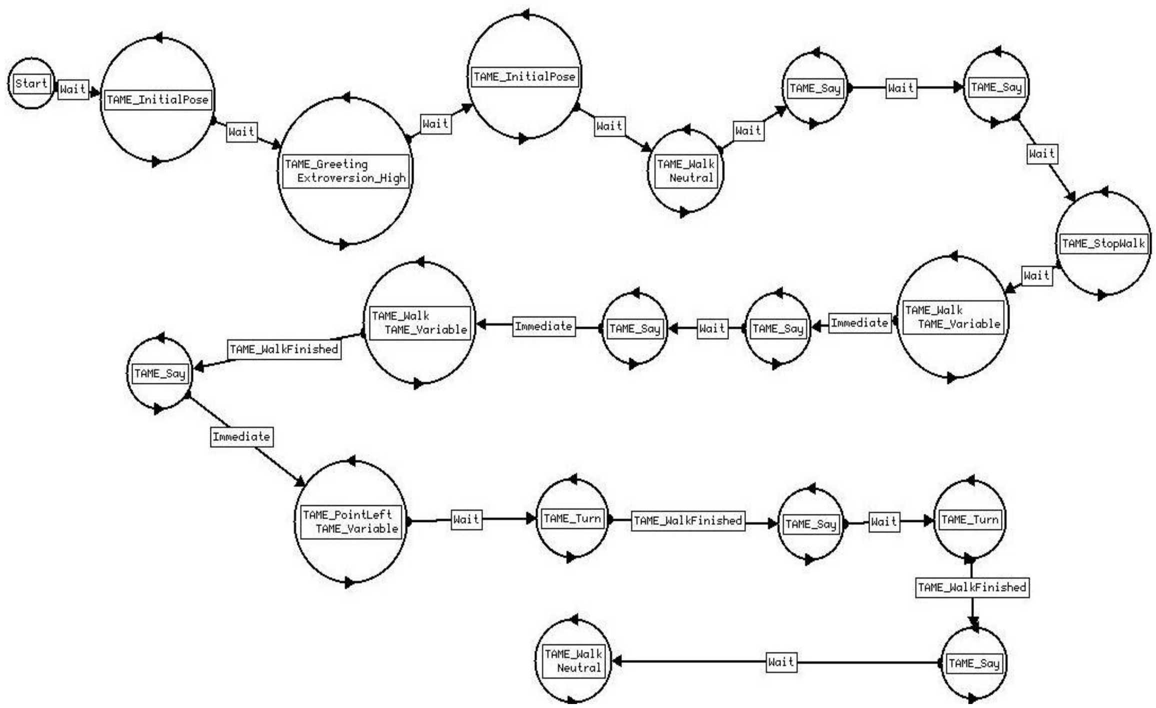


Figure 66: FSA used for sequencing Nao's behavior in the Search-And-Rescue experiment

The expressions of Negative Mood were almost identical to those used for the online recognition survey. To reduce the reported confusion with a more energetic and curious demeanor characteristic of Positive Affect, the robot's head was lowered even further, and the head turns to the side were faster and jerkier, to impart an impression of anxiety and vigilance, rather than interest and curiosity. The display of Fear was the same as in the survey (**Figure 67**). In both *Mood* and *Combined* conditions, the robot's walk towards the exit was faster than in the *Control* condition, to suggest a sense of urgency.

Figure 67: Expression of Fear on Nao robot

Additionally, paralinguistic cues were used to express Negative Affect and Fear through a synthesized male voice (speech was absent in the survey). For Negative Mood and Fear, the pitch was raised, and the rate of speech was increased [184] with the help of a TTS engine (CSLU toolkit [178]); the difference was more pronounced with Fear ("Something is wrong!" was the only phrase affected by fear, as emotions are short-lived). **Table 21** summarizes expressive differences between the conditions for each section of the robot's "tour".

Table 21: Expressive differences between the conditions, Mood experiment

Tour Section	Control	Mood	Combined
Before the overhead lights went out	A basic medium-speed walk with slightly swinging arms; default TTS voice; no differences between the conditions		
After the lights went out, but before pointing	No changes from the basic behaviors	Head lowered down; hands clenching/ unclenching and wrists turning left or right periodically; fast head movements to the left and to the right, as if checking what was going wrong; higher-pitched and faster speech.	Same as *Mood*

Table 21 (continued)

Saying "Something is wrong" and pointing	No changes	No bodily expressions; the speech is higher-pitched and faster than default.	After the pointing gesture, the robot crouched low to the ground, lowered its head down, and placed one hand in front of the face, as if covering it; the voice was higher-pitched and faster than in the Mood condition
Indirect and direct requests	No changes	No bodily expressions; the speech is higher-pitched and faster than default.	Same as *Mood*.
Walking towards the exit	No changes	Walking slightly faster than before	Same as *Mood*.

6.1.4 EXPERIMENT DESIGN

The goal of this experiment was to identify the effect of display of mood (negative) and emotions (fear) by a humanoid robot on participants' perception of the robot and on their compliance with the robot's request, in the context of a mock-up search-and-rescue scenario. The study followed 1-factor between-subject design with three conditions: *Control* (no affective expressions were displayed by the robot), *Negative Mood* (the robot displayed sings of Negative Affect), and *Combined* (the robot exhibited both Negative Mood and Fear). The between-subject design was necessitated by the fact that this study relied on the element of surprise (the participants did not expect to be ordered to evacuate), which would be lost in within-subject design.

6.1.4.1 Hypotheses

The following formed the hypotheses for this HRI experiment:

1. The interaction would be judged as more natural, persuasive and understandable in the *mood* and *combined* conditions, compared to *control*, and more so in the *combined* condition than in the *mood* condition.

2. The compliance with the robot's request would be the least in the *control* condition, followed by the *mood* condition, and the most in the *combined* condition. We hypothesized that the expression of robot's affective state would help the subjects assess the environment better and thus they would be more inclined to perform the robot's request.

3. The participants would experience greater Negative *Affect* in the *mood* and *combined* conditions than in *control*.

4. The participants would be able to recognize the display of *negative mood* and *fear* by the robot.

6.1.4.2 Measures

Evaluation was performed using both subjective (a variety of questionnaires) and objective (compliance) methods with respect to the aforementioned study hypotheses.

Hypothesis 1 was evaluated using three psychological scales designed for this experiment. Each scale consisted of 5 semantic differential subscales, where each subscale asked the respondent to indicate his/her position on a rating scale anchored by two bipolar words. The terms with negative connotation were placed on the left (at "1") of each 5-point subscale, and those with positive connotation on the right (at "5"); thus, higher scores signify a more positive report. The subscales within a scale were designed to measure the same overarching concept, as evidenced in the following scales used in this study:

- *Understandability* scale, designed to assess how understandable the robot's behavior was (**Figure 68**);

- *Persuasiveness* scale, designed to assess how persuasive the robot's request to leave was (**Figure 69**);

1. In your opinion, the robot's BEHAVIOR was:

Confusing				Clear
1	2	3	4	5

Unreadable				Easy to Read
1	2	3	4	5

Inconsistent				Consistent
1	2	3	4	5

Hard to Understand				Easy to Understand
1	2	3	4	5

Inexpressive				Expressive
1	2	3	4	5

Figure 68: *Understandability* Scale

1. In your opinion, the robot's REQUEST TO LEAVE was:

Ignorable				Compelling
1	2	3	4	5

Inappropriate				Appropriate
1	2	3	4	5

Ineffective				Persuasive
1	2	3	4	5

Insincere				Sincere
1	2	3	4	5

Unconvincing				Convincing
1	2	3	4	5

Figure 69: *Persuasiveness* Scale

- *Naturalness* scale, designed to assess how natural the robot appeared to the participants. This scale combines a number of subscales of two overlapping scales, *Anthropomorphism* and *Animacy*, presented in Bartneck et al. [80], and eliminates redundancy (**Figure 70**).

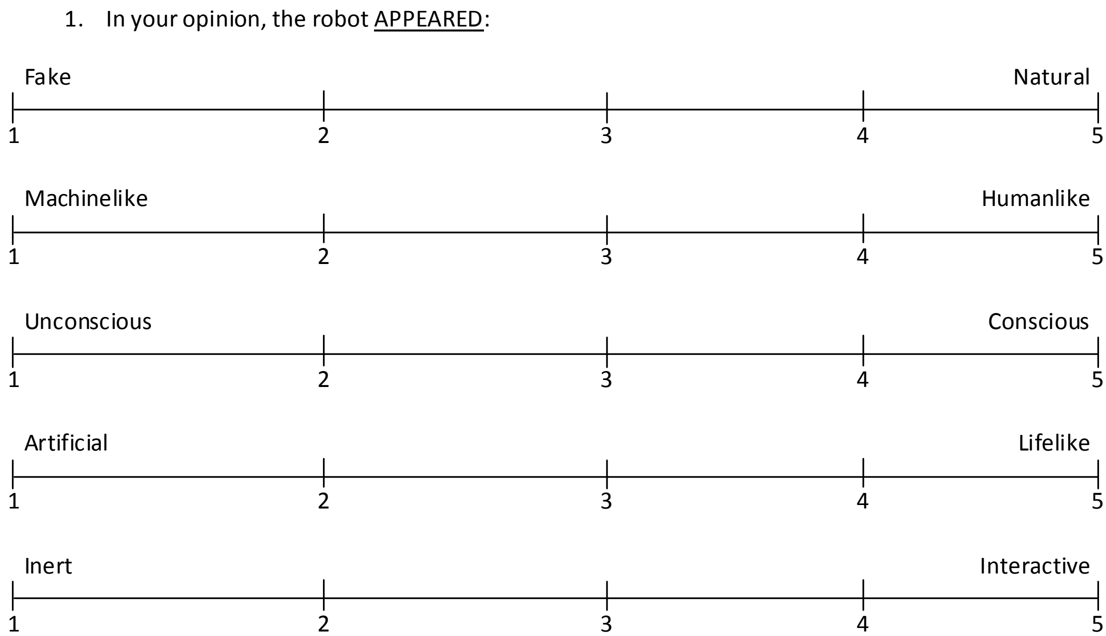


Figure 70: *Naturalness* Scale

Hypothesis 2 was evaluated by analyzing the video footage with respect to the following metrics:

- Whether a subject complied with robot's request by moving towards the exit, and the relative distance he/she traversed with respect to the cross markers (see **Figure 71** for location of the markers):
    - Moved to the first red cross marker, but not past;
    - Moved a little past the marker (subject still visible in the video);
    - Moved significantly past the marker (subject not visible in the video);

- How fast the subject traversed the distance (how much time elapsed between the robot's request to "proceed to the exit" and the subject's reaching the first marker);

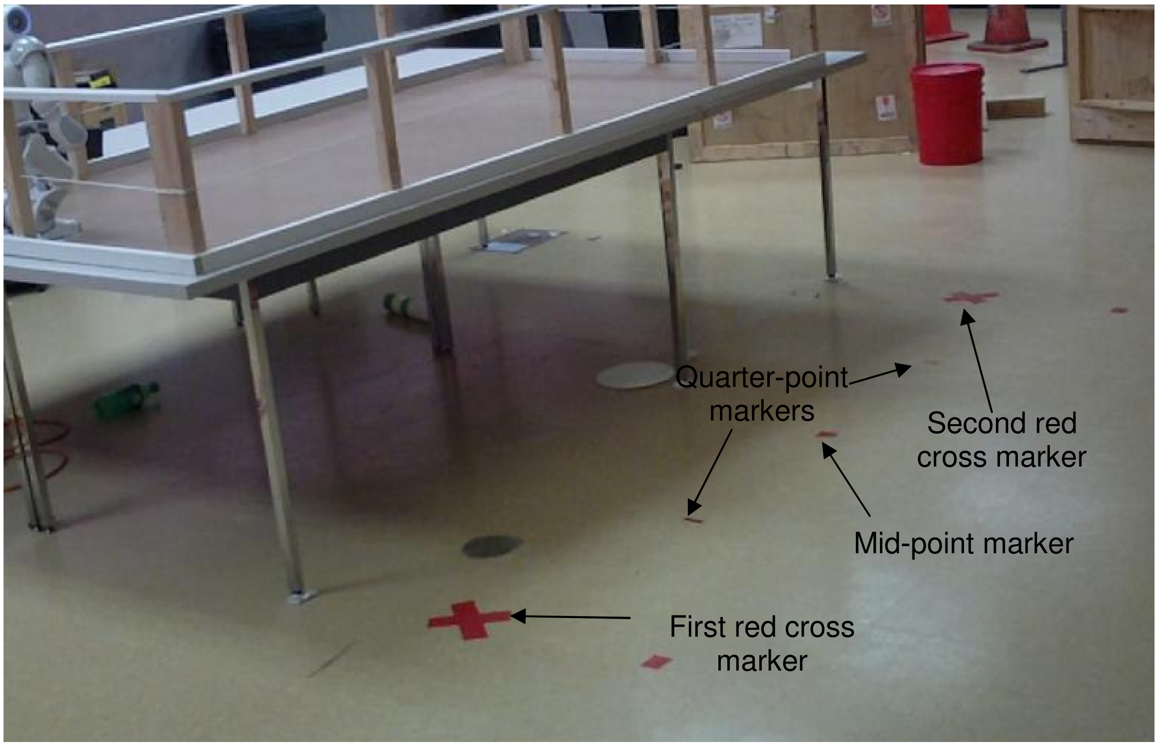


Figure 71: Location of the markers (from camera footage).

- How soon the participants reacted to the robot's request: in some cases, they moved a few steps in response to the indirect request, and reached the closest quarter-point marker or beyond (see **Figure 71** for marker location);

To assess the participants' mood (**Hypothesis 3**), an established psychological measure of mood PANAS-T (Positive/Negative Emotionality Schedule [162] was used (Appendix G), once before the experiment, and once right after the interaction part was over.

Finally, the following measures were used to assess robot affect recognition (**Hypothesis 4**):

- The same mood questionnaire, PANAS-T (Positive/Negative Emotionality Schedule [162] (Appendix S), used previously to assess participants' mood, only now with regards to the robot;
- A post-questionnaire question, requesting that participants specify whether the robot exhibited any of the following 6 emotions during the experiment: Anger, Joy, Fear, Disgust, Sadness and Interest, on a likert-style scale ranging from 1 (Not at all) to 5 (Extremely).

Additionally, the extent to which the robot's behavior affected the participants' decision to leave was measured by a 5-point likert-style scale, anchored with "Not at All" at 1 and "Extremely" at 5. This question, the three aforementioned scales and the emotion recognition question were combined together into a single post questionnaire (Appendix T). An open-ended question regarding the overall interaction with the robot concluded the post questionnaire which was followed by a short demographics questionnaire (Appendix U).

At the very end, the subjects were also requested to participate in a brief semi-structured interview, designed to uncover any misunderstandings, unexpected opinions and systematic inconsistencies (Appendix V).

6.1.5 ANALYSIS AND RESULTS

A total of 48 people participated in the experiment. The data for two of them were excluded from the analysis due to poor English and inability to understand the robot; one participant could not complete the experiment due to a technical problem. After an outlier analysis, data of two more participants were excluded due to either an overly positive or overly negative bias: the cumulative score on the post-questionnaire scales was outside of +/- 2 Standard Deviations of Mean. This left a total of 43 participants with valid questionnaire data, 14 each in *control* and *mood* conditions, and 15 in the *combined* condition.

6.1.5.1 Participant Demographics

The participants were recruited by two methods: 1) through Experimetrix, a GA Tech undergraduate psychology experiment pool (the students who completed the experiment were given ½ class credit for 30 minutes of participation), and 2) through flyers/word of mouth advertising (Appendix W) on GA Tech campus ($10 Starbucks gift cards served as compensation). Not surprisingly, the majority of the participants were undergraduate (60%) and graduate (21%) GA Tech students, in their 20s or younger (81%), or in their 30s (16%). In terms of gender composition, there were more males (60%) than females (40%), mostly equally distributed between the conditions (6 females each in *control* and *combined* conditions, and 5 in *mood*). The vast majority considered themselves either technical (67%), or somewhat technical (23%), and all of the participants were computer-savvy, at least at the level of advanced user. Finally, most of them had either no (28%) or limited (47%) robot experience, with only 6 participants having had experience with more than one robot type.

6.1.5.2 Hypothesis 1: Understandability, Persuasiveness and Naturalness

1-way ANOVAs were performed on the scores of each scale and subscale (*Understanding*, *Persuasiveness*, and *Naturalness*) in the post-questionnaire.

For *Understandability*, no significant differences at 0.05 level were observed either overall or for each individual subscale. Appendix X.1 provides descriptive statistics (number of cases, mean, Standard Deviation, and Standard Error) for *Understandability* construct. Overall, the participants found the robot's behavior during the experiment fairly understandable (average score of 18.33 out of 25), and the affective robotic expressions did not add to or subtract from this assessment. One interesting observation is that the subjects also did not find the robot's behavior in the affective conditions (*mood* and *combined*) any more expressive than in the *control* ($Mean_{control,\ expressive}$ = 3.1, $Mean_{mood,\ expressive}$ = 3, and $Mean_{combined,\ expressive}$ = 3.7, Appendix X.1).

For *Persuasiveness*, no significant differences at 0.05 level were observed overall; however, some differences on the subscale level were found (see Appendix X.2 for descriptive statistics).

In particular, the differences between the conditions in the scores of *Compelling* and *Sincere* subscales were significant at 0.05 level, with $F_{compelling} = 4.1$, $p<0.023$ and $F_{sincere} = 3.3$, $p<0.47$, respectively, and the difference between *Convincing* subscale scores was weakly significant at $F_{convincing} = 2.8$, $p<0.73$. Post-hoc comparisons (Least Significant Differences test, LSD hereafter) were performed on these subscales to identify more specific differences between the conditions. It was found that:

1. the robot in *Mood* and *Combined* conditions was perceived as more **compelling** than in *Control* ($p<0.013$ and $p<0.023$, respectively);
2. the robot in *Mood* condition was viewed as more **sincere** than in *Control* ($p<0.016$);
3. the robot in *Combined* condition was perceived as more **convincing** than in *Control* ($p<0.025$).

Figure 72-74 show Standard Error/Mean plots for *Compelling, Sincere* and Convincing subscales, respectively. Overall, the robot's request to leave was found to be rather persuasive (19.88 out of 25, as averaged between the 3 conditions).

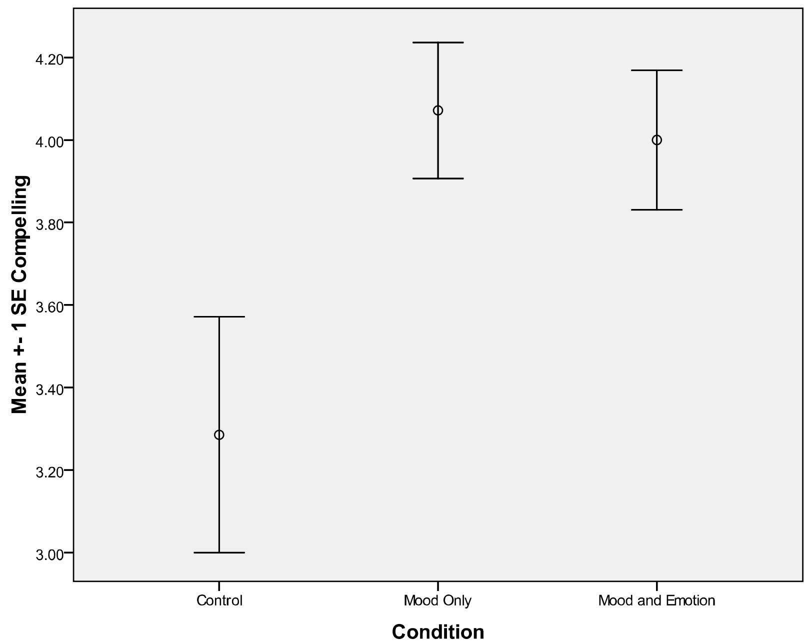


Figure 72: Standard Error/Means Plot for *Compelling* subscale: participants in *Mood* and *Combined* conditions found the robot's request the most compelling.

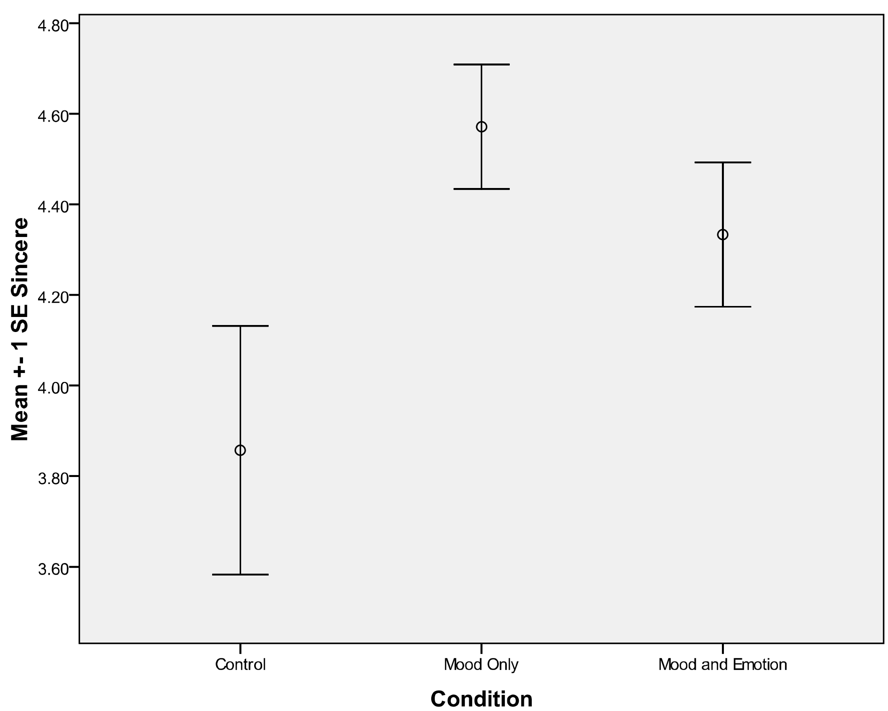


Figure 73: Standard Error/Means Plot for *Sincere* subscale: participants in *Mood* condition found the robot's request the most sincere.

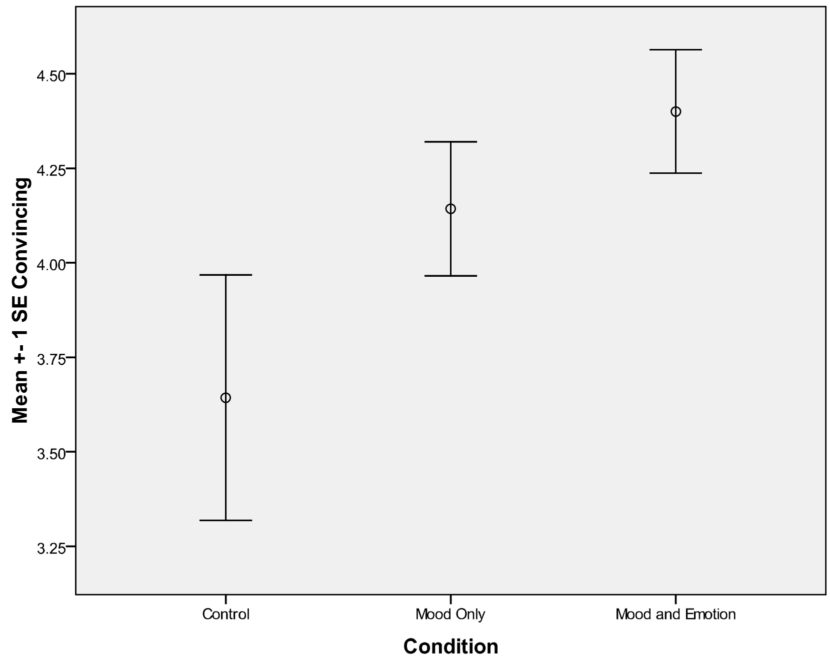


Figure 74: Standard Error/Means Plot for *Convincing* subscale: participants in *Combined* condition found the robot's request the most convincing.

Finally, for *Naturalness*, no significant differences at 0.05 level were observed overall; however, some differences on the subscale level were found (see Appendix X.3 for descriptive statistics).

In particular, the difference between the conditions in the scores for *Conscious* subscale was significant at 0.05 level: $F_{conscious} = 4.48$, $p<0.018$. Post-hoc comparisons (LSD) revealed that the robot in *Combined* condition appeared more **conscious** than in *Control*, $p<0.005$. **Figure 75** shows Standard Error/Mean plot for *Conscious* subscale. Overall, Nao in this experiment appeared neither natural nor unnatural (average score of 15.59 out of 25); however, the participants found it appeared particularly **conscious** (4.07 out of 5) in the *Combined* condition.

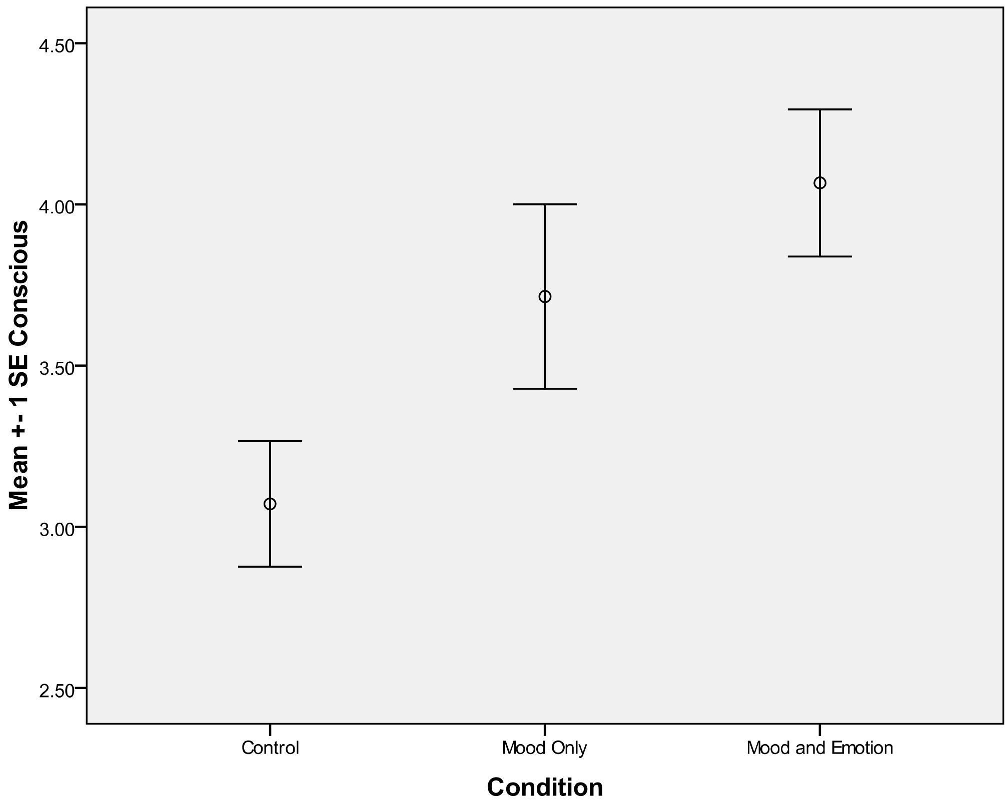


Figure 75: Standard Error/Means Plot for *Conscious* subscale: participants in *Combined* condition viewed the robot as more conscious than those in *Control*.

Additionally, Pearson's Correlation test revealed a significant correlation (2-tailed) at 0.001 level between the scales of *Persuasiveness* and *Naturalness* (see **Table 22** for results on all three scales), indicating that those who found the robot more natural also found its request to leave more persuasive. Interestingly, the same effect was not observed with *Understandability*, indicating that the participants did not connect their understanding of the robot's behavior with either persuasiveness of its request or naturalness of its appearance.

Table 22: Correlations between the scales of *Understandability*, *Persuasiveness* and *Naturalness: Persuasiveness* and *Naturalness* scales are strongly correlated

		Understandability	Persuasiveness	Naturalness
Understandability	Pearson Correlation	1	.097	.281
	Sig. (2-tailed)		.537	.068
	N	43	43	43
Persuasiveness	Pearson Correlation	.097	1	.412**
	Sig. (2-tailed)	.537		.006
	N	43	43	43
Naturalness	Pearson Correlation	.281	.412**	1
	Sig. (2-tailed)	.068	.006	
	N	43	43	43

**. Correlation is significant at the 0.01 level (2-tailed).

To summarize, although the respondents' understandability of the robot's behavior was not better in the affective conditions, the participants:

1. found the robot's request to leave more **compelling** and **sincere** in the *Mood* condition than in *Control*;
2. found the robot's request more **compelling** and **convincing** in the *Combined* condition than in *Control*;
3. rated the robot in the *Combined* condition as more **conscious** than the one in *Control*;
4. in general, found the robot more *Persuasive* when it appeared more *Natural*.

6.1.5.3 Hypotheses 2: Request Compliance

Video recordings were available for 14 participants in *Control* and *Mood* conditions each, and for 13 in *Combined* (due to technical difficulties, two sessions could not be filmed, and were treated as missing data). To evaluate whether the subjects complied with the robot's request better in the affective conditions, the available video recordings were analyzed in a number of ways:

1. Time elapsed (in seconds) between the robot's direct request to leave, "Proceed to the exit", and the moment of subject reaching the first cross marker was calculated, and 1-way ANOVA was performed on this variable (called "Time To Cross" from now on). This metric shows how fast the participants reacted to the robot's request; the marker was chosen as the end point because: 1) the subjects were already familiar with it (they were asked to stand on it in the beginning), 2) it signified the point beyond which the robot could not physically move, and 3) it was easily and reliably identifiable from video recordings. The ANOVA results were weakly significant, with $F_{TimeToCross}=2.61$, $p<0.09$ (see **Table 23** for means and SDs).

Table 23: Descriptive Statistics for *Time To Cross* variable

Dependent Variable	Condition	N	Mean	Std. Deviation	Std. Error
Time To Cross	Control	11	6.7273	2.32770	.70183
	Mood Only	12	5.4167	2.39159	.69039
	Mood and Emotion	11	4.0000	3.54965	1.07026
	Total	34	5.3824	2.93376	.50314

Post-hoc comparisons (LSD) revealed a significant difference between *Control* and *Combined* conditions, p<0.029, indicating that the participants in *Combined* condition took less time from the time the robot issued its direct request to leave until the time they reached the cross marker; **Figure 76** presents this result visually. Finally, given that our original hypothesis predicted that the participants in either of the affective conditions would be faster, a planned orthogonal comparison was performed, where the *Control* condition was compared against the average of both affective conditions (contrast coefficients used: -1; 0.5; 0.5). The result was significant at 0.05 level, $p<0.044$, indicating that together the affective conditions resulted in faster compliance. Overall, the participants took over a second less on average in *Mood* condition than *Control* (5.4 vs. 6.7, respectively), and over 2.5 seconds in *Combined* condition that *Control* (4 vs. 6.7, respectively) to reach the cross marker.

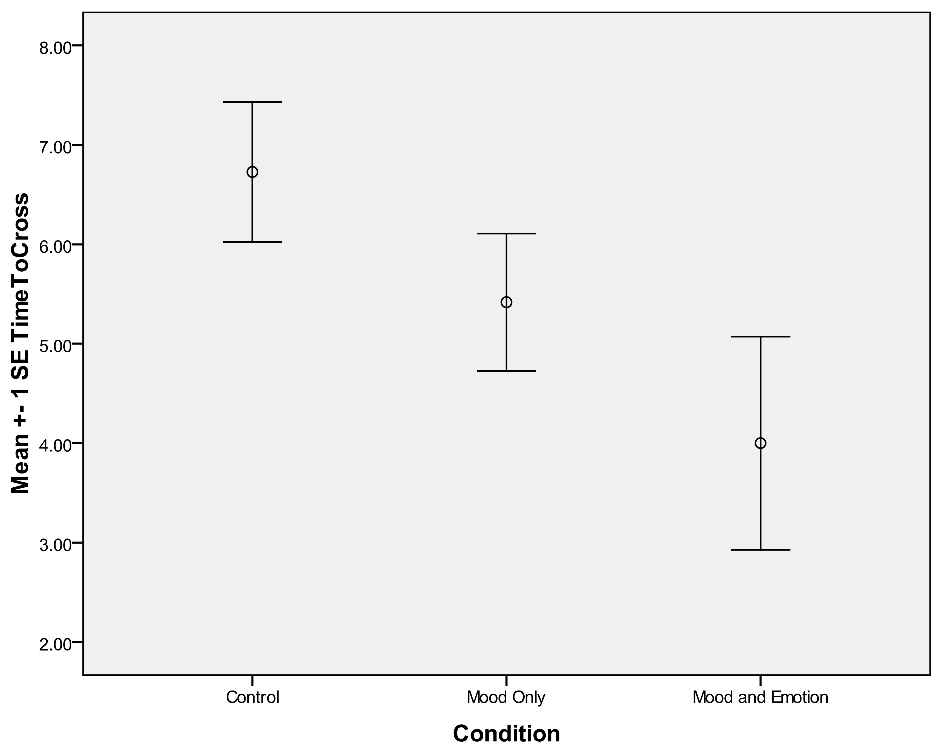


Figure 76: Standard Error/Means Plot for *Time To Cross* variable: participants in *Combined* condition took less time from the robot's request to proceed to the exit to reaching the first cross marker.

2. The relative distance the subjects traversed in response to the robot's request was determined from video recordings. The distance fell into one of four categories: 1) the participants did not move at all or did not reach the first cross marker ("No Walk"); 2) the participants stopped at the first cross marker ("At

Cross"); 3) the participants went a little past the marker, but were still visible in the video ("A Little Past"); 4) the participants moved fully outside of the camera view, and had to be stopped by the experimenter ("A Lot Past"). Due to wide difference in the participants' number in each of the categories, it was not possible to perform statistical analysis; however, **Figure 77** presents the differences between the conditions graphically. As we can see, only half (50%) of those in *Control* condition went past the cross marker, whereas in the affective conditions this percentage was higher: over 70% in *Mood* condition, and over 75% in *Combined*. This suggests that more participants in the affective conditions felt compelled to go further, thus complying with the request to a greater extent.

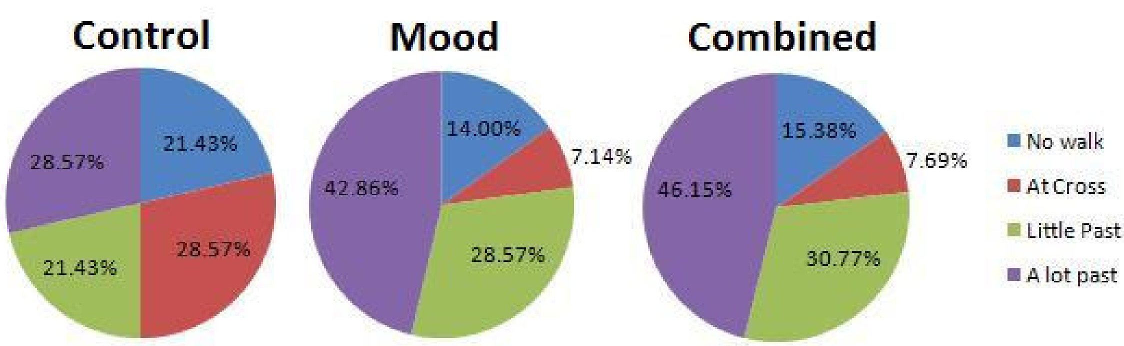


Figure 77: Percentage distribution of relative distance.

3. Finally, we noted at which point in time the participants started to react to the robot's requests. As you may recall, there were two separate requests made by the robot to the same end: in the first one, the robot uttered a less direct phrase "We need to evacuate immediately", and in the second one, the robot made a direct request: "Please proceed to the exit". We observed an interesting phenomenon: not a single participant in the *Control* condition reacted to the first (indirect) request, whereas 29% (4 out of 14) of those in the *Mood* condition and 31% (4 out of 13) of those in *Combined* took a few steps towards the exit after

the first (indirect) request, reaching or passing the closest quarter-point marker, and then stopped to wait for further instructions (most of them actually asked the robot what they should do next). This finding is displayed graphically in **Figure 78**. It appears, therefore, that the robot's expressive behavior in the affective conditions made the subjects more sensitive to the robot's message, more alert and eager to act even in response to an indirect request. It should be noted that the actions taken by the robot and the wording were exactly identical in all three conditions.

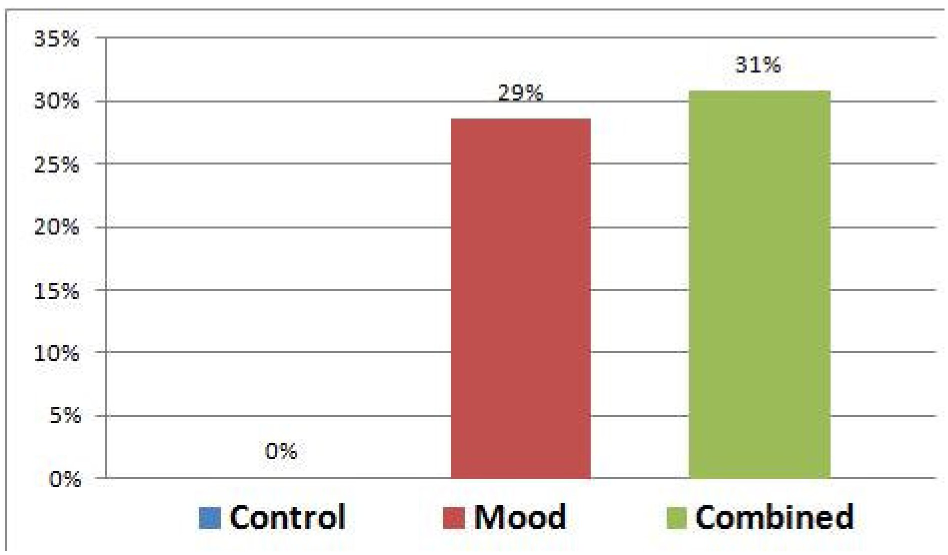


Figure 78: Percentage of participants who responded to robot's indirect request to evacuate: almost a third of the participants in the affective conditions responded to the indirect request, while none in the *Control* condition did.

To summarize, almost a third of the participants in the affective conditions seemed to respond to the robot's indirect request (with none in the *Control* condition); and a larger percentage of them were willing to go further when requested to leave, than those in *Control*. Also, the subjects in the *Combined* condition took less time between the robot's direct request and reaching the first marker than those in *Control*, suggesting a greater compliance, and even potential practical benefits, for example, in cases where mere time could make a difference between life and death.

6.1.5.4 Hypothesis 3: Participants' Negative Affect

It was hypothesized that the participants in the affective conditions will experience greater Negative Affect, as they might pick up on the anxiety and nervousness cues exhibited by the robot. To evaluate this hypothesis, a 1-way ANOVA was performed on the cumulative averaged Negative Affect score on PANAS-T test taken immediately after the interaction with Nao. The result was weakly significant ($F_{NegativeAffectAfter} = 2.58$, $p<0.081$), and LSD post-hoc comparison revealed that the participants reported a higher level of Negative Affect in *Combined* condition than in *Control* ($p<0.031$); see Appendix X.4 for descriptive statistics.

As PANAS-T includes multiple facets, some of which (e.g., depressed or hostile) are of no specific interest to this study, we also performed an ANOVA on more relevant facets. In particular, the difference between the scores of *Nervous* facet was significant at 0.05 level, with $F_{NervousAfter} = 4.71$, $p<0.015$; and post-hoc LSD comparison showed that the participants in the *Combined* condition felt more **nervous** than in *Control* ($p<0.004$) after they interacted with the robot (Appendix X.4 and **Figure 79**). Those in *Control* and *Mood* conditions reported feeling nervous at between "not at all" and "a little" level (with *Mood* condition reports closer to "a little": $Mean_{mood,\ NervousAfter} = 1.8$, and $Mean_{control,\ NervousAfter} = 1.5$), whereas those in *Combined* condition felt "a little" to "moderately" nervous ($Mean_{combined,\ NervousAfter} = 2.5$). No statistically significant differences were observed for scores of Negative Affect ($p<0.389$) and Nervousness ($p<0.107$) obtained as baseline mood ratings before the experiments (see Appendix X.4 for means and SDs), indicating that the differences were induced through the interaction with the robot.

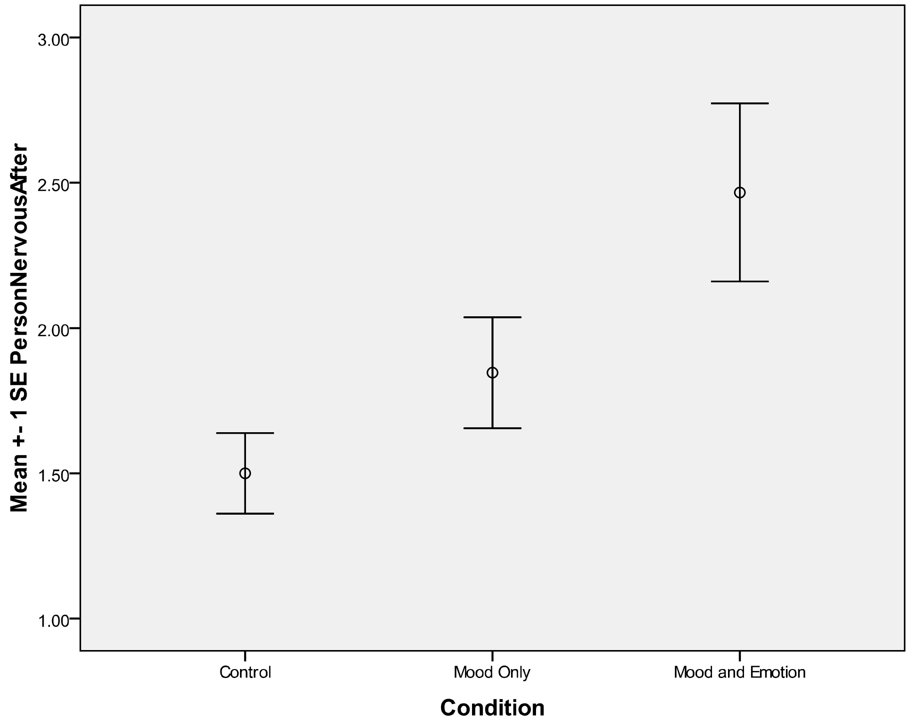


Figure 79: Standard Error/Means Plot for *Nervous* facet of PANAS-T test after their interaction with the robot: participants in the combined condition felt significantly more nervous than those in control.

Additionally, a 2x3 ANOVA was performed on Order (*Before* or *After* robot interaction) and Affect (*Control*, *Mood* and *Combined*) on *Nervous* facet; a significant main effect of Order ($F_{NervousOrder} = 7.4$, $p<0.008$) and a significant interaction effect ($F_{NervousInteraction} = 3.6$, $p<0.031$) were observed. These effects are displayed graphically in **Figure 80**; we can see a downward curve from *Control* to affective conditions before the experiment, and an upward curve after. A simple main effect of *Order* at *Combined* condition was also significant ($F_{NervousCombined} = 9.4$, $p<0.005$), showing that those in *Combined* condition became more **nervous** after their interaction with the robot.

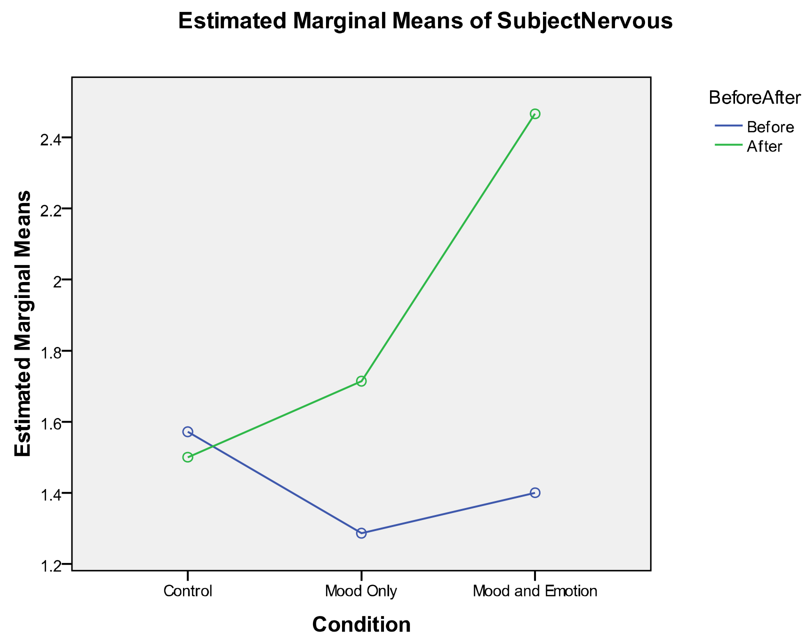


Figure 80: Means Plot of the *Nervous* scores by Order and Affect conditions. A main effect of order and an interaction effect were observed – the curve is downward from *Control* to affective conditions before the experiment, and upward after.

To summarize, the expressions of both Negative Mood and Fear on the robot did induce increased **nervousness** in the participants; however, the expressions of Negative Mood alone were quite not enough to do so.

6.1.5.5 Hypothesis 4: Robot's Negative Affect and Fear

To determine whether negative affect and fear in the robot were recognized by the participants, 1-way ANOVAs were performed on the average score for *Negative Affect* from PANAS-T questionnaire, and on *Fear* score from the post questionnaire. No statistically significant differences at the 0.05 level were found between the conditions, suggesting that the differences in the affect expression between the conditions were not consciously identified. **Table 24** presents descriptive statistics for these variables. Overall, the subjects believed that the robot experienced low levels of Negative Affect, and a small to moderate level of Fear (from 2.36 out of 5 in *Control* condition to 2.93 in *Combined*), but these levels did not differ significantly between the conditions.

Table 24: Descriptive Statistics for Robot Affect variables: participants reported low levels of robot's Negative Affect, and a small to moderate level of fear

Dependent Variable	Condition	N	Mean	Std. Deviation	Std. Error
Robot Negative Affect	Control	14	1.6104	.41034	.10967
	Mood Only	14	1.7662	.50768	.13568
	Mood and Emotion	15	1.8121	.51479	.13292
	Total	43	1.7315	.47733	.07279
Robot Fear	Control	14	2.3571	1.00821	.26945
	Mood Only	14	2.7143	1.43734	.38414
	Mood and Emotion	15	2.9333	1.27988	.33046
	Total	43	2.6744	1.24825	.19036

6.1.5.6 Other Observations

Although the following observations did not address the study hypotheses directly, they are nonetheless useful to obtain a more detailed picture of the results:

- Respondents' scores on question 5 of the post-questionnaire, "To what extent, if any, did the robot's behavior, and not its words, influenced YOUR DECISION to comply with the request to leave?", were strongly positively correlated with their scores on *Persuasiveness* scale at 0.01 level, and moderately negatively correlated with *Time to Exit* variable at 0.05 level, as evidenced by the results of Pearson's Correlation tests (**Table 25**). This indicates that those participants who thought the robot's behavior influenced their decision, also found the robot's request more persuasive and took less time between the robot's direct request to leave and reaching the first marker.

Table 25: Correlation results between *Decision To Leave*, *Time to Cross*, and *Persuasiveness* variables

		Decision To Leave	Time To Cross	Persuasiveness
Decision To Leave	Pearson Correlation	1	-.380*	.518**
	Sig. (2-tailed)		.027	.000
	N	43	34	43
Time To Cross	Pearson Correlation	-.380*	1	-.249
	Sig. (2-tailed)	.027		.155
	N	34	34	34
Persuasiveness	Pearson Correlation	.518**	-.249	1
	Sig. (2-tailed)	.000	.155	
	N	43	34	43

*. Correlation is significant at the 0.05 level (2-tailed).

**. Correlation is significant at the 0.01 level (2-tailed).

- A significant correlation at 0.05 level was observed between the levels of participants' Negative Affect **after** their interaction with the robot and their ratings of robot's Negative Affect. Similarly, there was a correlation at 0.05 level between how Nervous the participants felt **after** their interaction, and how they rated the robot on its overall Negative Affect, and the nervous subscale (**Table 26**). This indicates a link between subjects' assessment of their own affective state, and of the robot's; it may mean that either they are projecting their own feelings onto the robot, or that the robot's behavior was instrumental in inducing a greater level of Negative Affect and nervousness in humans. It should also be noted that there were no significant correlations between the ratings on participant's *Negative Affect* and *nervousness* **before** their interaction with the robot, and corresponding robot affect.

Table 26: Correlation between participants' and robot's *Negative Affect*: there was a correlation between how nervous subjects reported feeling after their interaction with the robot, and their reports of robot's Negative Affect and nervousness

		Subject Negative Affect, After	Subject Nervous, After	Robot Negative Affect	Robot Nervous
Subject Negative Affect, After	Pearson Correlation	1	.830**	.366*	.267
	Sig. (2-tailed)		.000	.017	.087
	N	42	42	42	42
Subject Nervous, After	Pearson Correlation	.830**	1	.348*	.389*
	Sig. (2-tailed)	.000		.024	.011
	N	42	42	42	42
Robot Negative Affect	Pearson Correlation	.366*	.348*	1	.715**
	Sig. (2-tailed)	.017	.024		.000
	N	42	42	43	
Robot Nervous	Pearson Correlation	.267	.389*	.715**	
	Sig. (2-tailed)	.087	.011	.000	
	N	42	42	43	

6.1.6 SUMMARY AND DISCUSSION

The goal of this experiment was to determine whether robotic expressions of mood and emotions – Negative Mood and Fear in this specific case – may provide identifiable benefits for human-robot interaction. The first three of the hypotheses posited a priori were in regards to this goal, and were for the most part validated in the results of the study (with the exception of the subjects' ratings of the robot's *Understandability*, where no significant differences between the conditions were observed). In particular:

- the participants found the robot's request to evacuate more **compelling**, **sincere** and **convincing** in one or both affective conditions than in control;
- they complied with the robot's request to "evacuate" to a greater extent in the affective conditions:

  - the subjects were **faster** in complying with the robot's request to leave the "dangerous" zone (in the Combined condition);
  - they were **more prone** to respond to an **indirect request** to evacuate in both of the affective conditions;
  - **more** of those in the affective conditions **walked further** towards the exit than in the control.
- the participants reported feeling more nervous after interacting with the robot in the Combined condition than in control, potentially making them more alert to any unfavorable changing in the surroundings.

These were the expected results; the rest of the subsection will focus on the findings that were less straightforward.

1. **Understandability.** It was hypothesized that the participants would find the robot's behavior more understandable in the affective conditions, but this did not happen. One potential reason for why this was the case could be as follows: as understanding is a conscious cognitive process, and affect recognition occurs on a more automatic, even visceral level, the finding that the participants could not consciously identify the expressions of negative mood and fear in the robot could have contributed to their not finding the robot's behavior more understandable. Another possible explanation behind this finding lies, perhaps, in the perceived dissonance between the robot's acting anxious and the subjects' assurances regarding the safety of the environment – it was just an experiment, after all.
2. **Negative Mood and Fear recognition.** The lack of differences between conditions in subjects' ratings of robot's Negative Affect and Fear was rather

unexpected, given the positive results of a prior affect recognition survey. There could be a number of explanations for this discrepancy:

- In the survey, the respondents were attending to the robot's behavior only, whereas in the search-and-rescue experiment, the robot's behavior was a part of a more complex situation, and the attention was split between the mock-up accident site setup, the information provided by the robot's speech, robot's actions, all this making any expressions the robot exhibited much less salient.
- Within-subject design was used in the survey, to reduce individual differences. This design was not possible for the search-and-rescue experiment, as the participants would not respond in the same manner if they have knowledge of a future evacuation request. However, it would stand to reason that a within-subject design would increase the recognition: during pilot testing, a third of the participants were showed one of the other conditions after they completed one, and most of the time, they could identify the differences.

3. **Negative Mood Only vs. Mood + Fear.** Perhaps not surprisingly, the results for the *Combined* condition were overall stronger. It was predicted that the response would be less pronounced in the *Mood* only condition, but, unfortunately, in many cases the differences between *Mood* and *Control*, and *Mood* and *Combined* conditions did not reach statistical significance, even though they were in a predicted direction. Given the subtle nature of mood expressions, it is possible that the power would have been increased with a greater number of participants.

6.2 EVALUATING EXPRESSIONS OF EXTRAVERSION AND INTROVERSION IN A ROBOT AS A GUIDE SCENARIO

Different personality traits allow people to adjust to different environments, tasks and jobs they perform in life. In particular, it was found that Extraverts are better suited for jobs that require gregariousness and energy, such as teaching and leadership [185-187]. People are also very good at recognizing nonverbal expressions of Extraversion, even based on short exposure [183], as well as are capable of correctly differentiating between an extraverted and introverted humanoid robot from video clips [165]. However, the same characteristics that would be beneficial to a task requiring engagement and gregariousness (e.g., teaching or guiding), may be detrimental to a task that calls for concentration from its participants (e.g., problem solving). Therefore, the trait of Extraversion (at two ends of the continuum, high and low) was chosen for evaluation in the second HRI experiment due to its potential practical applicability and proven recognizable characteristics, with the goal to determine whether different personality manifestations in a humanoid robot are more suitable for different types of tasks. There were a total of two between-subject experimental conditions: *Extraverted* (the robot behaved in a manner characteristic of highly extraverted individuals) and *Introverted* (the robot behaved in a manner characteristic of one low in extraversion).

6.2.1 EXPERIMENT SCENARIO AND SETUP

In this experiment, a humanoid robot (Nao) played the role of a guide at a mock-up building demolition exhibit. The "guided tour" consisted of two parts: 1) a robot presentation on the building demolition process followed by a subject-matter quiz taken by the participants; and 2) robot's supervision of a problem solving task. Prior to robot interaction, the participants performed similar baseline tasks to both familiarize themselves with the types of tasks expected of them and to establish a baseline for comparison (to reduce the impact of individual differences). The entire experiment, from greeting to compensation, lasted 45-60 minutes, and the interaction with the robot took approximately 10 minutes.

The experiment took place in the same lab as the search-and-rescue study, but the setup was modified to resemble a building demolition exhibit. In particular, a number of posters with images of explosive building demolition were placed around, to provide a museum-like setting. The setup is shown in **Figure 81**. The same small humanoid robot (Aldebaran Robotics' Nao) was used, and placed in the same arena as in the previous experiment, to avoid any accidental falls. The robot was set facing the participants, who were instructed to stand in front of the robot at a marker for the duration of the interaction.

Figure 81: Mock-up Building Demolition exhibit setup.

6.2.2 EXPERIMENT PROCEDURE

This subsection describes the experimental procedure, from greeting a subject to providing compensation for participation. After greeting the subject at the entrance to the building where the lab was located, he/she was first taken to the portion of the lab which was separated from the experiment setup. There, the participants were asked to read over and sign a consent form (Appendix Y) and a video release form (Appendix P). Those without an equivalent of at least high school education and English language proficiency were excluded from this study (the task in the study required a good command of English and a basic knowledge of algebra). The subjects were then told that they would perform two baseline tasks to prepare them for similar tasks during their interaction with the robot; they were also asked to remember that the tasks were not a competition or a race, and that they should work at their own pace (to put them at ease). The tasks presented to the subjects were as follows:

1. ***Memory Retention (Quiz)***. For this task, the participants listened to a presentation given either by a male undergraduate GA Tech student in a video recording (baseline task), or by the robot (experimental task). The subject matter between the presentations was slightly different, but comparable in difficulty and length and within the same general area (building construction practices for the baseline text, and explosive building demolition for the experimental one). After the presentation, the subjects took a short quiz consisting of 5 multiple-choice questions; they were instructed to answer based on the material in the presentation, and not any general knowledge they may have. Both presentations and quizzes were pilot-tested previously to make sure the difficulty was appropriate, and that the two tasks (baseline and experimental) were comparable. Please see Appendix Z.1 for the baseline presentation and quiz, and Appendix Z.2 for the experimental presentation and quiz (2 presentation versions: *Extraverted* and *Introverted*). The presentations took approximately 6-7 minutes, and answering the quizzes took 84 seconds on average (as timed by the experimenter).

2. ***Problem Solving (Math).*** For this task, the subjects were asked to solve a simple math problem. The experimental problem was taken from a Kaplan SAT practice math test (multiple-choice), and was rated of medium difficulty. It was then modified for wording but not for numbers to fit the subject matter of the presentation, to create a more consistent experience for the participants. The baseline problem was similar, but both numbers and wording were modified to insure enough differentiation between the problems. In the baseline case, the experimenter gave the problem instructions to the participants, and was present during the problem solving (timing the duration). The experimental task was a variant of a dual-task problem [188], where the participants were expected to perform two things simultaneously: solve the math problem, and attend to the robot. During this task, although it was the experimenter who provided the physical sheet of paper with the problem on it, the robot gave the instructions, and the participants were asked to follow the robot's instructions carefully, thus constantly attending to the robot. There was a time limit of 180 seconds on the experimental problem. The baseline math problem can be viewed in Appendix AA.1, and robot's instructions (*Extraverted* and *Introverted* versions) and the experimental problem in Appendix AA.2.

After both baseline tasks were completed, the participants took a brief personality test [163], while the experimenter prepared the robot behind the partition. After they were done, the participants were asked to imagine they were at a fun building demolition exhibit in an industrial museum where a humanoid robot would serve as their guide. They were then were taken to the mock-up exhibit setting through a separate entrance. Once inside, the subjects were instructed to stand in front of the robot (at a marker to maintain consistency) and to follow its instructions carefully.

Both tasks (presentation and math) were preceded by a short (under a minute) greeting from the robot, introducing itself and the exhibit (see Appendix BB for *Extraverted* and *Introverted* greeting). This allowed the participants to familiarize themselves with the robot and its artificial speech. At the end of the greeting, Nao encouraged the participants to look around the “exhibit”, and to return when they were done. After each task, a series of task-specific questionnaires was given to the subjects, and an overall post-questionnaire, a personality questionnaire with regards to the robot, and a demographics questionnaire were given after both tasks were completed. Once all the questionnaires were filled out, the experimenter thanked the participants and provided the compensation.

6.2.3 ROBOTIC IMPLEMENTATION

Aldebaran Robotics’ humanoid robot Nao was used for this experiment; the behaviors and personality expressions were programmed as described in Chapter 5, *Software Architecture and Implementation*, using *MissionLab* and the integrated *TAME* module. The same nonverbal characteristics of Extraverted and Introverted individuals were taken into account while programming the expressions of Extraversion and Introversion as described in subsection 5.3.1.1, *Nonverbal personality display*. However, additional gestures and static poses were designed and implemented on the robot due to a longer duration of the robot’s performance than what was used for the affect recognition survey.

In addition to gestures and body posture characteristic of this personality trait (hunched, close to the body, low energy for introverts, and erect, expansive, and energetic for extroverts), there was also a difference in the speed with which these gestures and posture shifts were performed: faster and more forceful on *Extraverted* Nao, and slower on *Introverted*. Finally, in terms of differences in interpersonal distance, the *Introverted* robot was placed 6” (1.5 times) farther away from the edge of the platform than *Extraverted*. **Figure 82** and **83** show initial poses for *Introverted* and *Extroverted* Nao, respectively, as well as relative distance from the participants (the camera was positioned in the same spot). **Figure 84** gives an example of a shy introverted pose, with hands brought together at the torso level, and **Figure 85** shows, from left to right, a progression of a chopping gesture used by *Extraverted* Nao for emphasis, from the highest point (chest level) to the lowest.

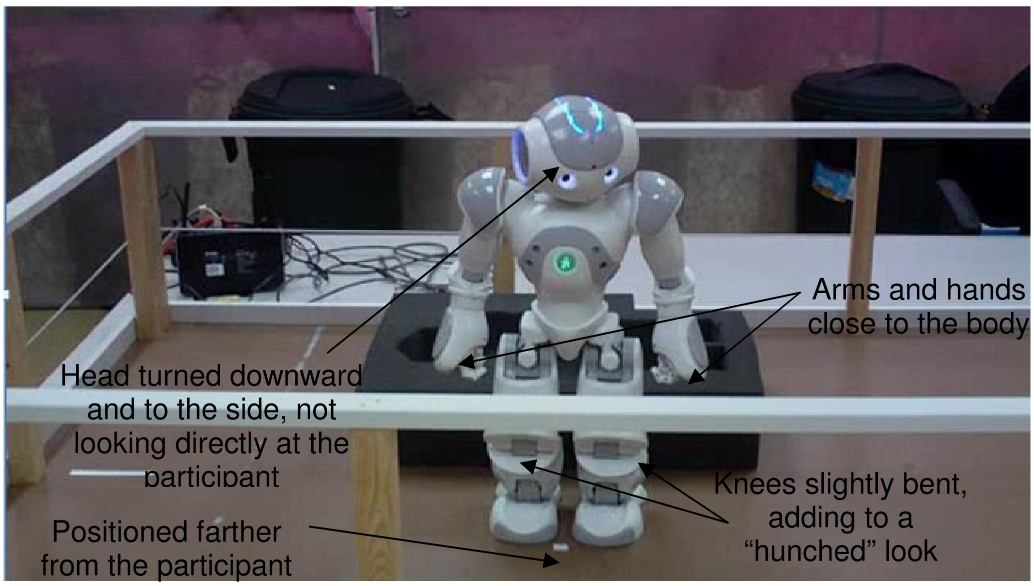


Figure 82: Initial static pose for Introverted Nao

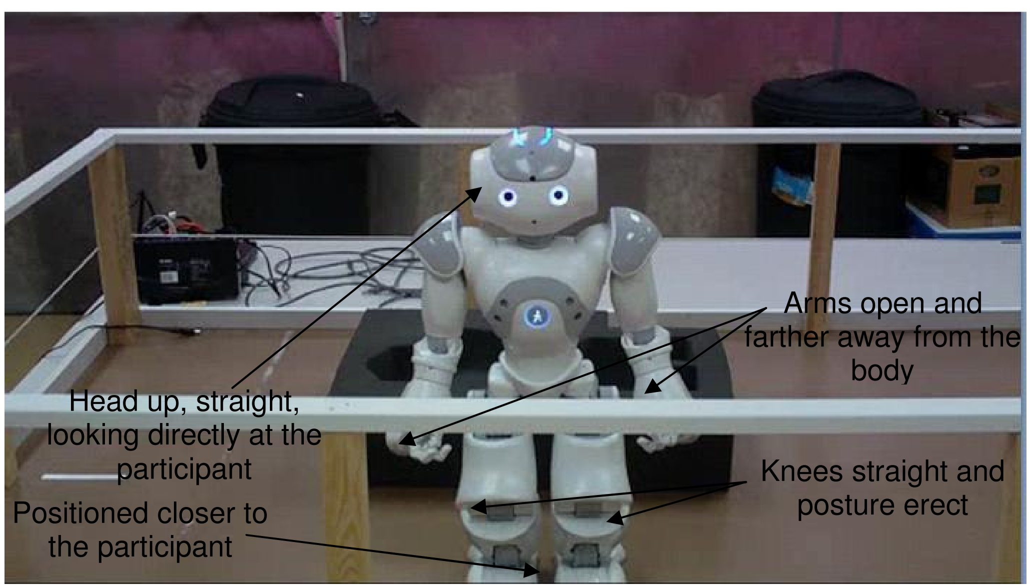


Figure 83: Initial static pose for Extraverted Nao.

Figure 84: Introverted Nao's shy pose

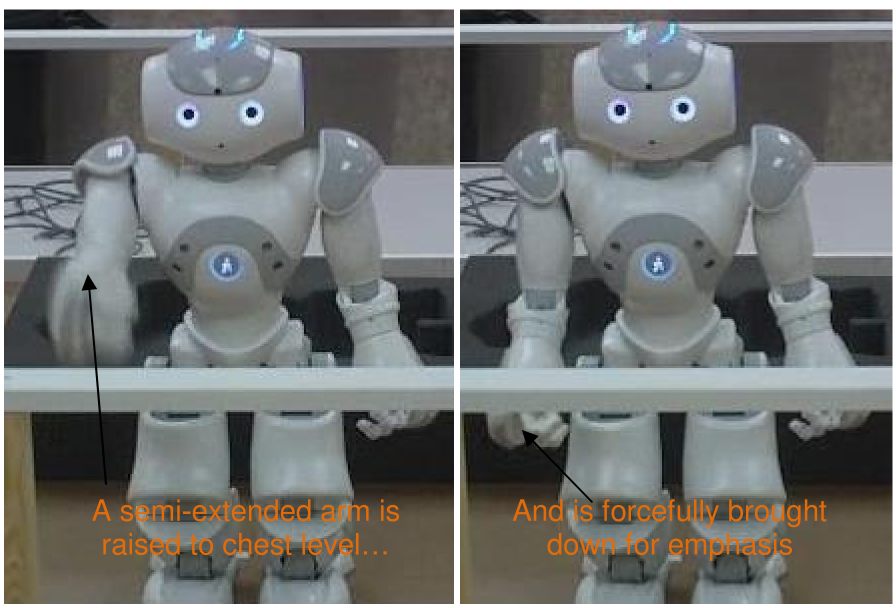


Figure 85: Progression of an emphatic chopping gesture on extraverted Nao, from the highest point to the lowest (from left to right).

The sequences of poses and gestures were scripted in an FSA and manually matched to the speech. There was a difference in frequency of gesturing and posture shifts during greeting, presentation and math instructions: the *Extraverted* robot performed a gesture or changed its pose every 1-6 seconds, whereas the *Introverted* one every 4 – 15 seconds. During the math task, after the participants started working on the problem, *Extraverted* Nao moved (including taking several steps along the arena and back), gestured or uttered an engaging phrase (Appendix AA) every 8-20 seconds, whereas *Introverted* Nao did so only occasionally (every 30-40 seconds), and did not engage in any walking behaviors, only posture shifts.

A TTS engine (CSLU toolkit [178]) was used to record Nao's speech on a male voice, and to encode paralinguistic cues: *Extraverted* Nao's speech was faster, louder, higher-pitched and more varied in pitch than *Introverted* Nao's. Additionally, short phrases or words were inserted into *Extraverted* robot's text for emphasis (e.g., "isn't it exciting?", "boom-boom", "I've been waiting for you", etc.), but factual information remained the same for both personalities (Appendix Y). Extensive pilot testing was done to ensure that the speech was equally intelligible in both cases, and easily identifiable differences were present (e.g., the faster the speech the harder it is to understand, so a balance had to be found). **Table 27** shows expressive differences between conditions.

Table 27: Expressive differences between *Extraverted* and *Introverted* conditions

Expression Type	Extraverted Robot	Introverted Robot
Posture	Erect, legs straight, head up and looking directly at the participants, arms away from the body	Hunched over, legs slightly bent, head down and to the side, arms close to the body
Gestures	Expansive, exaggerated, fast and forceful	Narrow, subdued, slow and weak
Movement Frequency	Frequent and varied	Infrequent and monotonous
Speech	Loud, fast, high-pitched and with wide pitch variability	Quieter (but loud enough to be heard distinctly), slow, low-pitched and with narrow pitch variability.

6.2.4 EXPERIMENT DESIGN

The goal of this experiment was to identify the effect of *Extraverted* and *Introverted personality* display by a humanoid robot on participants' task performance (to establish whether some traits are task-appropriate) and their perception of robot's appropriateness, friendliness, intrusiveness and naturalness in the context of a mock-up building demolition scenario. The study followed a 1-factor between-subject design with two conditions: *Extraverted* and *Introverted*, where the display of Extraversion or Introversion served as independent variables. Two experimental tasks were performed by participants in both conditions, with one task hypothesized to be better suited for an *Extraverted* robot, and the other for an *Introverted* robot. The experimental tasks were counterbalanced for order in both conditions; the baseline tasks were performed in the same order as the experimental tasks.

6.2.4.1 Hypotheses

The following formed the hypotheses for this study:

1. The *Extraverted* robot would be judged as more friendly (welcoming) and appropriate, and the facts presented by it as more appealing during the *quiz* task;
2. The *Introverted* robot would be judged as more appropriate and less obtrusive during the *math* task, and the task itself would appear easier in this condition.
3. The performance on the *math* task (correctness, completion time and reduction in completion time between baseline and experimental tasks) would be greater in the *Introverted* condition, as we hypothesized that the robot would provide fewer distractions and less annoyance, thus suggesting that a task requiring concentration would be more congruent with an introverted personality of the companion. Vice versa, the performance on the *quiz* task would be greater in the

Extraverted condition, in accordance with findings on correlation between extraversion and teaching and leadership effectiveness [185-187].

4. The participants would be able to recognize the display of *Extraversion and Introversion* in the robot.

6.2.4.2 Measures

Evaluation was performed using both subjective (a variety of questionnaires) and objective (task performance) methods with respect to the aforementioned study hypotheses.

Hypothesis 1 was evaluated using three psychological scales designed for this experiment, structurally similar to those in the previous experiment. Each scale consisted of 5 semantic differential subscales, where each subscale asked the respondent to indicate his/her position on a rating scale anchored by two bipolar words; the terms with negative connotation are placed on the left (at 1) of each 5-point scale, and those with positive on the right (at 5). The subscales within a scale were designed to measure the same overarching concept, as evidenced in the following scales used in this study:

- *Appropriateness* scale, designed to measure how well the robot's behavior matched the task it was performing; this scale was used for both experimental tasks, *quiz* and *math* (**Figure 86**);
- *Welcome* scale, designed to determine how welcome the robot made the participants feel (**Figure 87**);
- *Appeal* scale, designed to identify how appealing the participants found the facts presented by the robot (**Figure 88**).

Together, the three scales formed a post-quiz questionnaire (Appendix CC).

1. In your opinion, FOR THIS TASK, the robot's behavior was:

Inappropriate — Appropriate
1 2 3 4 5

Wrong for Task — Right for Task
1 2 3 4 5

Ill-suited — Well-suited
1 2 3 4 5

Improper — Proper
1 2 3 4 5

Mismatched — Matched to Task
1 2 3 4 5

Figure 86: *Appropriateness* scale

1. In your opinion, YOUR PRESENCE during the interaction with the robot was:

Unwelcome — Welcome
1 2 3 4 5

Undesired — Desirable
1 2 3 4 5

Disliked — Liked
1 2 3 4 5

Tolerated — Encouraged
1 2 3 4 5

Unwanted — Wanted
1 2 3 4 5

Figure 87: *Welcome* scale

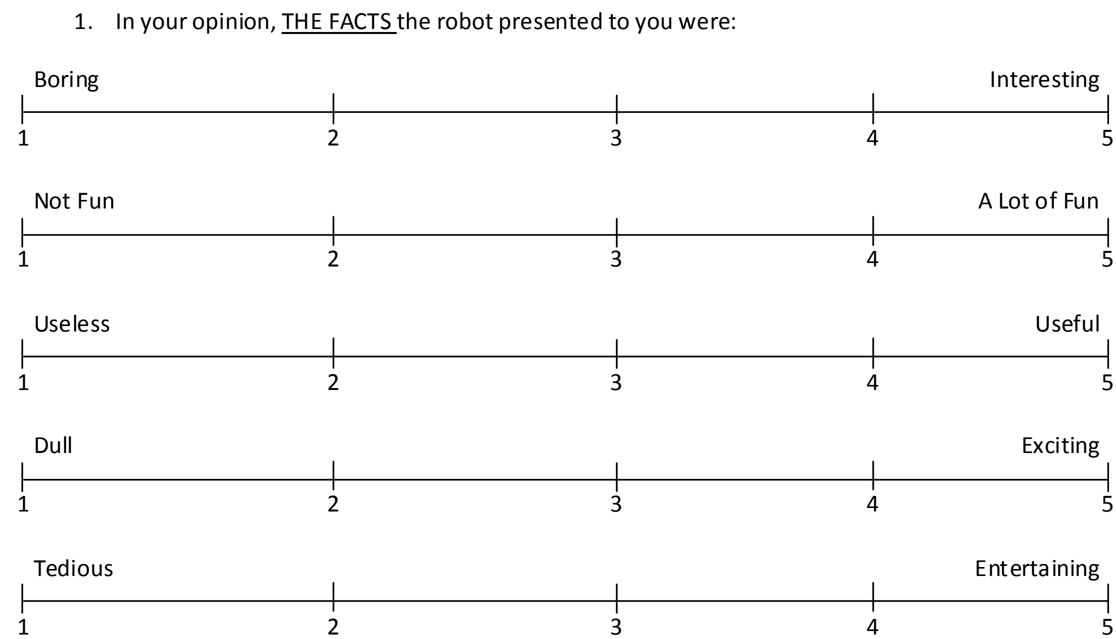


Figure 88: *Appeal* scale

To evaluate **Hypothesis 2**, the *Appropriateness* semantic differential scale (**Figure 86**) was used, along with two other scales:

- *Unobtrusiveness* scale, designed to measure how distracting the robot was during the *math* task. In this scale, the **higher** the score, the **less** distracting (or more unobtrusive) was the robot (**Figure 89**);
- *Ease* scale, designed to identify how easy the math problem was perceived to be (**Figure 90**).

Together, *Appropriateness*, *Unobtrusiveness* and *Ease* scales formed a post-math questionnaire (Appendix DD).

1. In your opinion, during this task, the ROBOT WAS:

Distracting				Easy to Tune Out
1	2	3	4	5

Interfering				Minding its own business
1	2	3	4	5

Annoying				Inoffensive
1	2	3	4	5

Irritating				Undemanding
1	2	3	4	5

Bothersome				Quiet
1	2	3	4	5

Figure 89: Unobtrusiveness scale

1. The MATH PROBLEM you have just solved was:

Hard				Easy
1	2	3	4	5

Complicated				Simple
1	2	3	4	5

Demanding				Undemanding
1	2	3	4	5

Long				Short
1	2	3	4	5

Complex				Basic
1	2	3	4	5

Figure 90: Ease scale

In addition, an established multidimensional measurement tool, Task Load Index (NASA-TLX [189], Appendix EE) was used to assess the perceived difficulty/workload of the *math* task. It provides an overall workload score based on a weighted average of ratings on six 21-point subscales: mental demand, physical demand, temporal demand, performance, effort and frustration. The reason for using this tool was two-fold: 1) to obtain greater detail and more insights on subjects' perceptions of the task demands, and 2) to compare the results of this more involved and time-consuming tool with a much shorter *Ease* scale, especially when used on such a simple task as a SAT-level math problem.

To evaluate **Hypothesis 3**, the quizzes and math problems (both experimental and baseline), were graded for correctness: "correct" or "incorrect" for math, and an overall cumulative score for each of the quizzes, with one point given for each correct answer. The experimenter also timed how long it took the participants to compete each task, from the moment they were told: "You can start now" till they announced they were done. Some practice effects were observed in the math task, so in addition to completion time, a percentage of improvement over the baseline completion time was also calculated.

Hypothesis 4 was assessed by analyzing the Extraversion scores of the brief version of Goldberg's Unipolar Big-Five Markers questionnaire (Mini-Markers) [163] with regards to the robot (Appendix FF).

Finally, the participants were also asked to fill out a post-questionnaire after both tasks were completed (Appendix GG). It contained a slightly modified *Naturalness* scale (**Figure 70**), where "inert/interactive" word pair was replaced with "inanimate/animate", and two open-ended questions, asking to compare the robot's behavior and subjects' perceptions between the two tasks, as well as to compare their prior expectations of robots with the impressions from the experiment. As usual, a demographics questionnaire was given at the end of the session (Appendix HH).

6.2.5 ANALYSIS AND RESULTS

There were a total of 30 participants in this study, 15 per condition. In one session (*Extraverted* condition), the robot malfunctioned during the *math* task (after the *quiz* task was completed), and the *math* task data from this session were treated as missing. In another session (*Introverted* condition), a participant was not able to complete the math task within the time limit. As this was the only participant not being able to complete the task within the allotted time, and given that the time limit (180 seconds) was outside of +/- 2 SD of the mean, the *math* task data were excluded from the analysis as an outlier. Therefore, the data of 15 participants in each condition were available for analysis for the *quiz* task, and the data of 14 participants in each condition for the *math* task. Additionally, one participant (*Extraverted* condition) completed the *TLX* questionnaire incorrectly, and these data were treated as missing for *TLX* score calculation.

To identify any task order effects, 2(Order) x 2(Trait) ANOVAs were performed on all dependent variables. No significant effects due to Order (*quiz* first vs. *math* first) were found, with two notable exceptions which are described in detail in subsection 6.2.5.2, *Hypothesis 1: Appropriateness, Welcome and Appeal in quiz task* and subsection 6.2.5.4, *Hypothesis 3: Task Performance*.

6.2.5.1 Participants' Demographics

The study participants were recruited by two methods: 1) through Experimetrix, a GA Tech undergraduate psychology experiment pool (the students who completed the experiment were given 1 class credit for 60 minutes of participation), and 2) through flyers/word of mouth advertising on GA Tech campus ($15 Starbucks gift cards served as compensation). Not surprisingly, the vast majority of the participants were undergraduate GA Tech students (87%), in their 20s (63%) or younger, but over 18 (33%). In terms of gender composition, there was the same number of women as men, equally distributed between the conditions. The vast majority considered themselves either technical (50%), or somewhat technical (40%), and all of the participants were computer-savvy, at least at a level of advanced user (with 1 exception). Most of them had either no (23%) or limited (43%) robot experience, with only 4 participants having had experience with more than one robot type. Finally, for 20% of the participants English was not a native language; however, they did have the required English language proficiency.

6.2.5.2 Hypothesis 1: Appropriateness, Welcome and Appeal in quiz task

It was hypothesized that for the *quiz* task, the participants in the *Extraverted* condition would find the robot more appropriate and welcoming, and the presentation facts more appealing. To analyze this hypothesis, 2- and 1-tailed (where appropriate) independent samples T-tests were performed on *Appropriateness*, *Welcome* and *Appeal* scales, as well as on the corresponding subscales.

For the *Quiz Appropriateness* scale, no statistically significant differences were found at 2-tailed level either for the overall scale or for the subscales (see **Table 28** for descriptive statistics). However, the differences between the conditions on *Appropriate* and *Well-suited* subscales were significant at 1-tailed level, with $t_{appropriate} = 1.77$, $p<0.044$, and $t_{well\text{-}suited} = 1.97$, $p<0.029$. This indicates that the participants found the *Extraverted* robot more **appropriate** and **well-suited** for the *quiz* task (giving a presentation) than the *Introverted* robot. Overall, the participants found the *Extraverted* robot largely appropriate (average score of 20.53 out of 25) for the *quiz* task, and the *Introverted* robot moderately so (18.4 out of 25).

Table 28: Descriptive statistics for *Appropriateness* scale with regards to *quiz* task

	Condition	N	Mean	Std. Deviation	Std. Error Mean
Appropriate (quiz)	Extraverted	15	4.3333	.61721	.15936
	Introverted	15	3.7333	1.16292	.30026

Table 28 (continued)

Right for Task (quiz)	Extraverted	15	4.0000	.84515	.21822
	Introverted	15	3.5333	1.12546	.29059
Well-suited (quiz)	Extraverted	15	4.2000	.77460	.20000
	Introverted	15	3.5333	1.06010	.27372
Proper (quiz)	Extraverted	15	4.0667	.88372	.22817
	Introverted	15	4.0000	1.06904	.27603
Matched to Task (quiz)	Extraverted	15	3.9333	.88372	.22817
	Introverted	15	3.6000	1.29835	.33523
Appropriateness (quiz), Overall	Extraverted	15	20.5333	3.41983	.88300
	Introverted	15	18.4000	5.01142	1.29394

Interesting Order (*quiz* first vs. *math* first) effects were found on the *Appropriateness* variable. A 2x2 (Order x Trait) ANOVA showed a significant interaction effect ($F_{interaction} = 7.66$, $p<0.01$) and a weakly significant main effect of Order ($F_{order} = 4.11$, $p<0.053$). **Figure 91** displays the interaction effect graphically: those who completed the *math* task first, found the *introverted* robot less **appropriate** (for the subsequent *quiz* task) than the *extraverted*, but those who were given the *quiz* task first, found the two personalities almost equally appropriate (**Table 29**).

Table 29: Descriptive statistics for *Quiz Appropriateness* for *math first* or *quiz first* task order

	Condition	N	Mean	Std. Deviation	Std. Error Mean
Appropriateness, ***math first***	Extraverted	8	21.0000	2.72554	.96362
	Introverted	8	15.3750	4.03334	1.42600
Appropriateness, ***quiz first***	Extraverted	7	20.0000	4.24264	1.60357
	Introverted	7	21.8571	3.67099	1.38750

This finding is confirmed by a significant simple main effect of *Trait* on *Appropriateness* for *math first*: $t_{appropriatenessAtmathFirst} = 3.27$, $p<0.006$; the difference for *quiz* first was not significant. Please note that it was the ratings of *Introverted* robot that changed so drastically between different task orders. One potential explanation for this curious finding could be as follows: it is possible that the *Introverted* robot's apparent suitability for the math task made its shortcomings during the presentation more obvious.

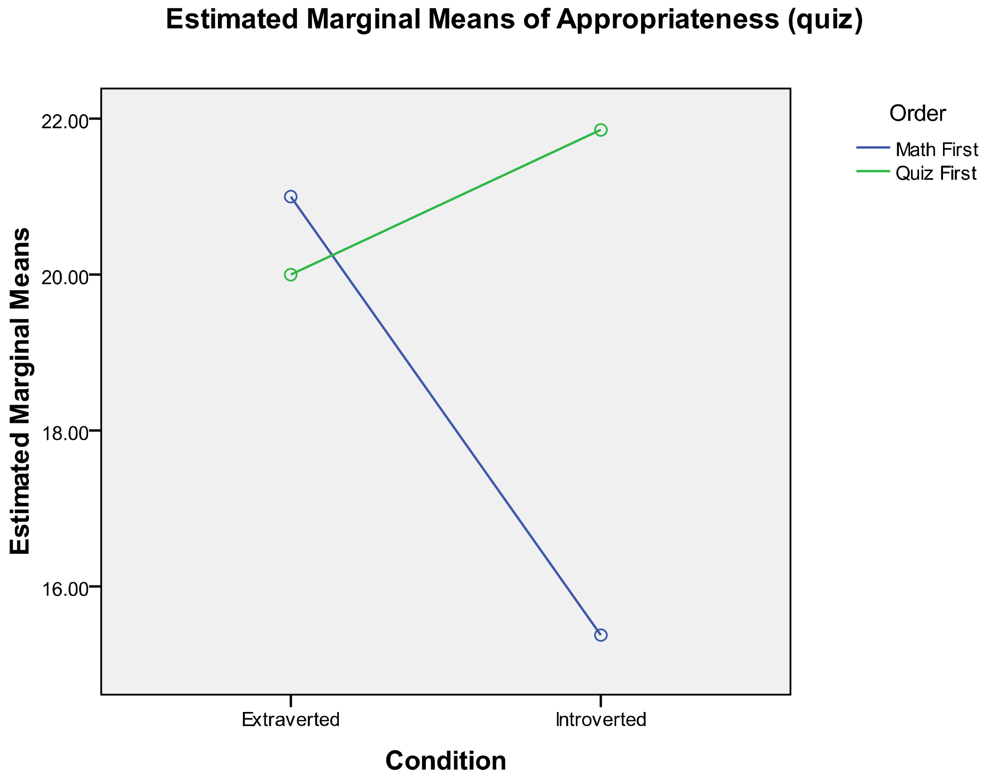


Figure 91: Trait x Order Means plot; in *math first* order, the *introverted* robot was found less appropriate for *quiz* task.

For *Welcome* scale, a 2-tailed independent samples T-test was significant at 0.01 level in the predicted direction, with $t_{Welcome} = 3.3$, $p<0.003$ (**Figure 92** presents this result graphically). The results for each subscale were significant as well, at either 0.01 or 0.05 level, with the exception of *welcome* subscale (see **Table 30** for descriptive statistics).

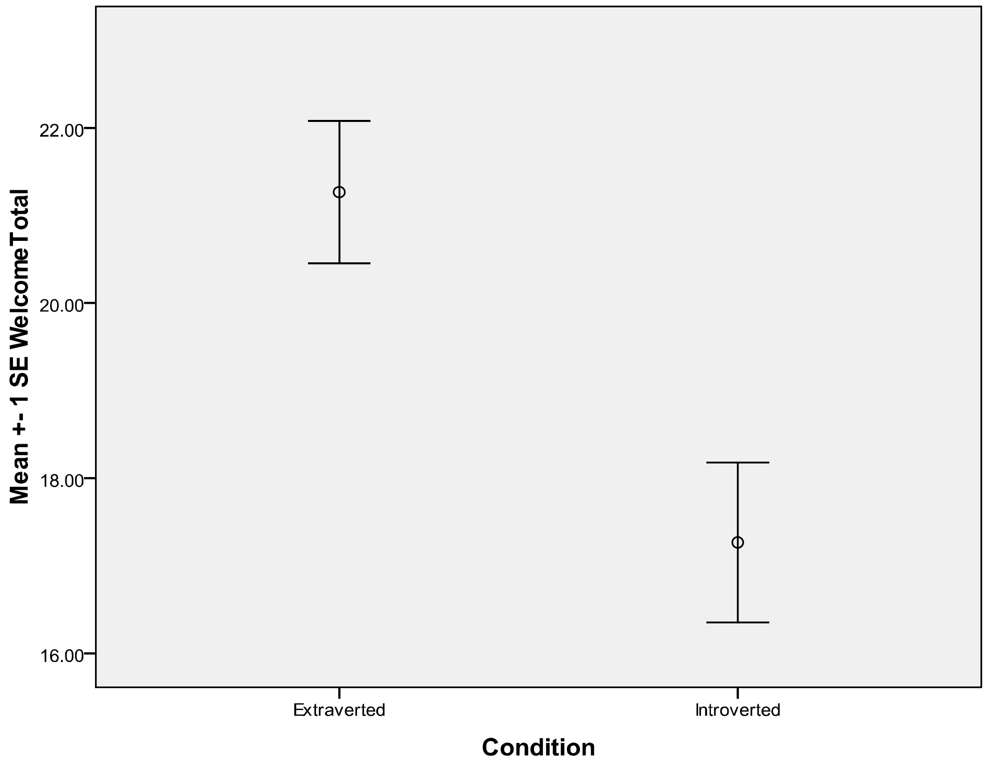


Figure 92: Standard Error/Means Plot for *Welcome* scale: participants in *Extraverted* condition felt more welcome overall.

Table 30: Descriptive statistics for *Welcome* scale

	Condition	N	Mean	Std. Deviation	Std. Error Mean
Welcome	Extraverted	15	4.2667	.79881	.20625
	Introverted	15	3.8000	.94112	.24300
Desirable	Extraverted	15	4.2667	.70373	.18170
	Introverted	15	3.2667	.79881	.20625
Liked	Extraverted	15	4.2667	.70373	.18170
	Introverted	15	3.2667	.70373	.18170
Encouraged	Extraverted	15	4.4000	.63246	.16330
	Introverted	15	3.5333	.99043	.25573
Wanted	Extraverted	15	4.0667	.88372	.22817
	Introverted	15	3.4000	.82808	.21381
Welcome, Overall	Extraverted	15	21.2667	3.15021	.81338
	Introverted	15	17.2667	3.53486	.91270

For Appeal scale, the result of a 2-tailed independent samples t-test was significant at 0.05 level in the predicted direction, with $t_{Appeal} = 2.7$, $p<0.012$ (see **Figure 93** for the Standard Error/Means plot, and **Table 31** for descriptive statistics). On the subscale level, the differences between the conditions for *Fun* and *Exciting* ratings were significant at 2-tailed level ($t_{fun} = 2.7$, $p<0.011$, and $t_{exciting} = 3.2$, $p<0.003$), and for *Interesting* at 1-tailed level ($t_{Interesting} = 1.9$, $p<0.033$). This suggests that the participants perceived the **same** explosive building demolition facts as being more **fun**, **exciting**, **interesting**, and **appealing** (overall) when they were presented by the *Extraverted* robot.

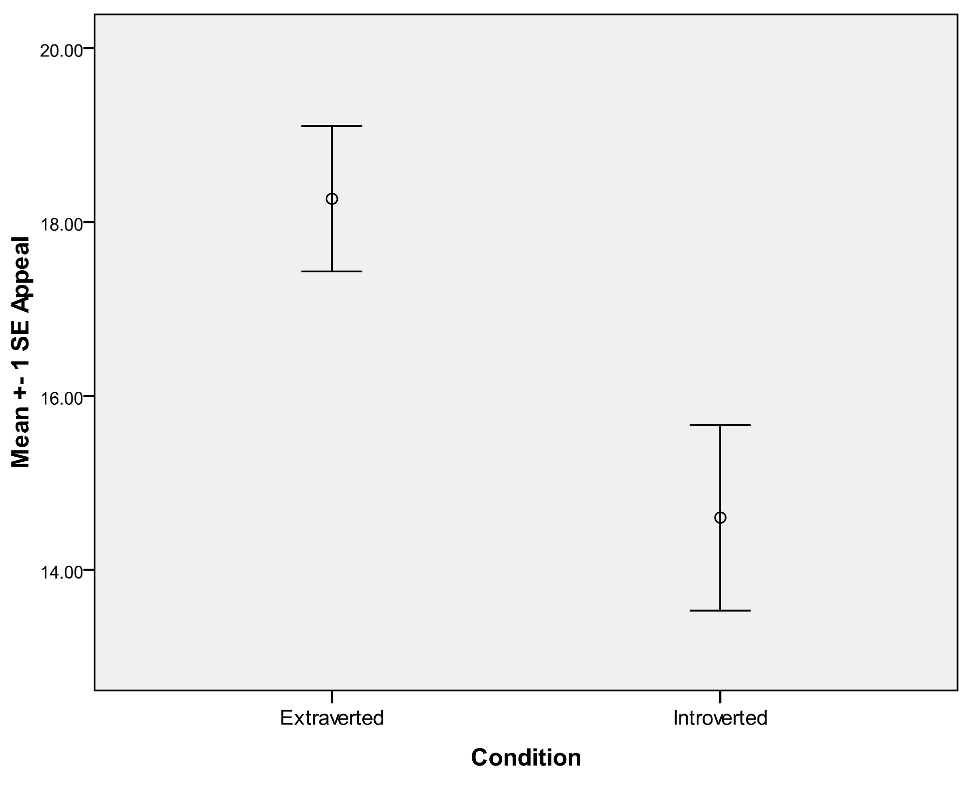


Figure 93: Standard Error/Means Plot for *Appeal* scale: participants in *Extraverted* condition found the facts presented by the robot more appealing

Table 31: Descriptive Statistics for *Appeal* scale

	Condition	N	Mean	Std. Deviation	Std. Error Mean
Interesting	Extraverted	15	4.2667	.79881	.20625
	Introverted	15	3.4667	1.40746	.36341
Fun	Extraverted	15	3.4667	.74322	.19190
	Introverted	15	2.6000	.98561	.25448
Useful	Extraverted	15	3.8667	1.06010	.27372
	Introverted	15	3.2667	1.16292	.30026
Exciting	Extraverted	15	3.4667	.74322	.19190
	Introverted	15	2.5333	.83381	.21529
Entertaining	Extraverted	15	3.2000	.77460	.20000
	Introverted	15	2.7333	1.09978	.28396
Appeal, Overall	Extraverted	15	18.2667	3.23964	.83647
	Introverted	15	14.6000	4.13694	1.06815

Additionally, Pearson's Correlations test revealed a significant correlation (2-tailed) at 0.001 level between the scales of *Welcome* and *Appeal* (see

Table 32 for results on all three scales), suggesting that those who felt more **welcome** during the presentation also found the facts more **appealing**. The scales of *Appropriateness* and *Appeal* were also found to be moderately correlated at the 0.05 level, thus linking the perceptions of task appropriateness with presentation appeal.

To summarize, it was found that, as predicted:

- the participants reported that the *Extraverted* robot in the quiz task made them feel as if their presence during the tour was more **liked**, **encouraged**, **wanted** and **desirable**;

- the **facts** presented by the *Extraverted* robot appeared more **fun**, **exciting** and **interesting** than those given by the *Introverted* robot.

Additionally, those who participated in the *math* task first, also found the *Extraverted* robot more appropriate for the quiz task, as compared to its introverted counterpart.

Table 32: Correlations between *Quiz Appropriateness*, *Welcome* and *Appeal* scales: those participants who felt welcome (overall) also found the facts given by the robot more appealing

		Appropriateness (quiz)	Welcome	Appeal
Appropriateness (quiz)	Pearson Correlation	1	.238	.419*
	Sig. (2-tailed)		.205	.021
	N	30	30	30
Welcome	Pearson Correlation	.238	1	.508**
	Sig. (2-tailed)	.205		.004
	N	30	30	30
Appeal	Pearson Correlation	.419*	.508**	1
	Sig. (2-tailed)	.021	.004	
	N	30	30	30

*. Correlation is significant at the 0.05 level (2-tailed).

**. Correlation is significant at the 0.01 level (2-tailed).

6.2.5.3 Hypothesis 2: Appropriateness, Unobtrusiveness and Ease in math task.

It was hypothesized that for the *math* task, the participants will find the Introverted robot more appropriate and unobtrusive, and the math problem easier. To analyze this hypothesis, 2- and 1-tailed (where appropriate) independent samples T-tests were performed on *Appropriateness*, *Unobtrusiveness*, *Appeal*, and *TLX* scales, as well as on the corresponding subscales.

The results of 2-tailed independent-samples T-tests on *Appropriateness* scale with respect to *math* task and on all its subscales were statistically significant at 0.01 level, with $t_{MathAppropriateness} = -4.2$, $p<0.0001$ (see **Table 33** for descriptive statistics). As hypothesized, the *Introverted* robot was found more appropriate for the *math* task than its extraverted counterpart; **Figure 94** presents this result visually for the overall scale.

Table 33: Descriptive statistics for *Appropriateness* scale with respect to *math* task

	Condition	N	Mean	Std. Deviation	Std. Error Mean
Appropriate (math)	Extraverted	14	2.7857	1.36880	.36583
	Introverted	14	4.2857	.72627	.19410
Right for Task (math)	Extraverted	14	2.5714	1.15787	.30945
	Introverted	14	4.2143	.80178	.21429
Well-suited (math)	Extraverted	14	2.7143	1.13873	.30434
	Introverted	14	4.0000	1.03775	.27735
Proper (math)	Extraverted	14	2.5714	1.15787	.30945
	Introverted	14	4.3571	.84190	.22501
Matched to Task (math)	Extraverted	14	2.4286	1.15787	.30945
	Introverted	14	3.7857	.89258	.23855
Appropriateness (math), Overall	Extraverted	14	13.0714	5.66316	1.51354
	Introverted	14	20.6429	3.58645	.95852

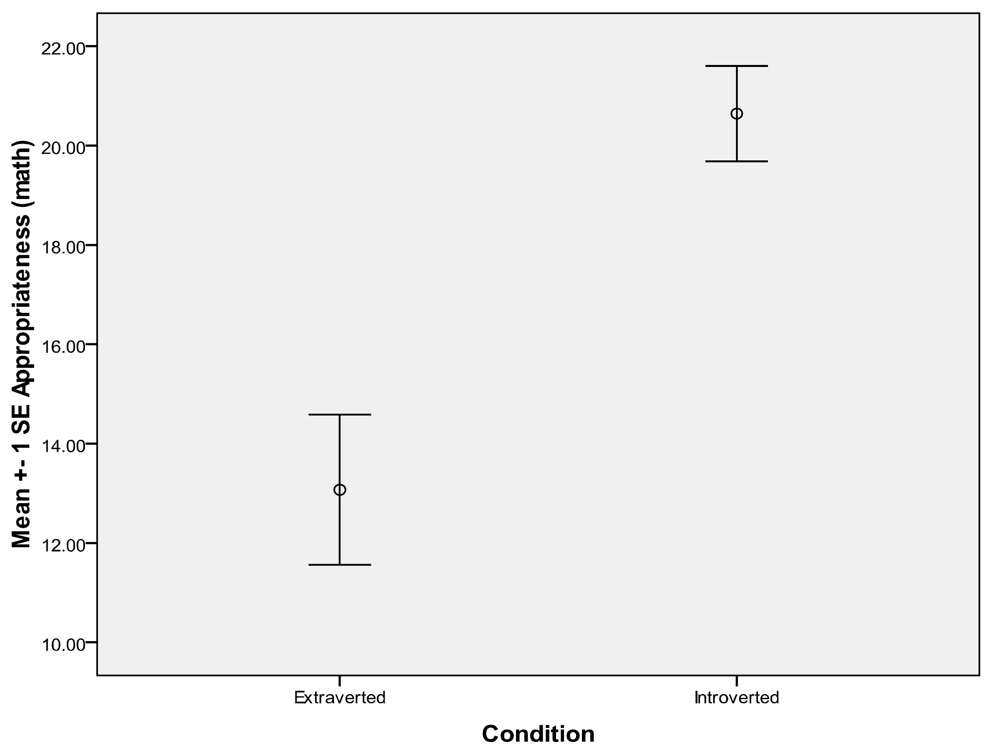


Figure 94: Standard Error/Means Plot for *Appropriateness* scale: the participants found the *Introverted* robot more appropriate for the *math* task.

The results on *Unobtrusiveness* scale and **all** its subscales were similarly significant at 0.01 level in the predicted direction, with $t_{Unobtrusiveness} = -5.4$, $p<0.000$ (see **Table 34** for descriptive statistics). This finding shows that the *Introverted* robot was perceived as less obtrusive than the *Extraverted* one (**Figure 95**): it was more inoffensive, undemanding and quiet, easier to tune out, and minding its own business.

Table 34: Descriptive Statistics for *Unobtrusiveness* scale

	Condition	N	Mean	Std. Deviation	Std. Error Mean
Easy to Tune Out	Extraverted	14	2.5000	1.34450	.35933
	Introverted	14	4.5714	.51355	.13725
Minding its Own Business	Extraverted	14	2.3571	.84190	.22501
	Introverted	14	4.1429	.94926	.25370
Inoffensive	Extraverted	14	2.8571	1.02711	.27451
	Introverted	14	4.1429	.94926	.25370
Undemanding	Extraverted	14	3.0714	1.07161	.28640
	Introverted	14	4.2857	.82542	.22060
Quiet	Extraverted	14	2.7143	.91387	.24424
	Introverted	14	4.0714	.82874	.22149
Unobtrusiveness, Overall	Extraverted	14	13.5000	4.23811	1.13268
	Introverted	14	21.2143	3.26234	.87190

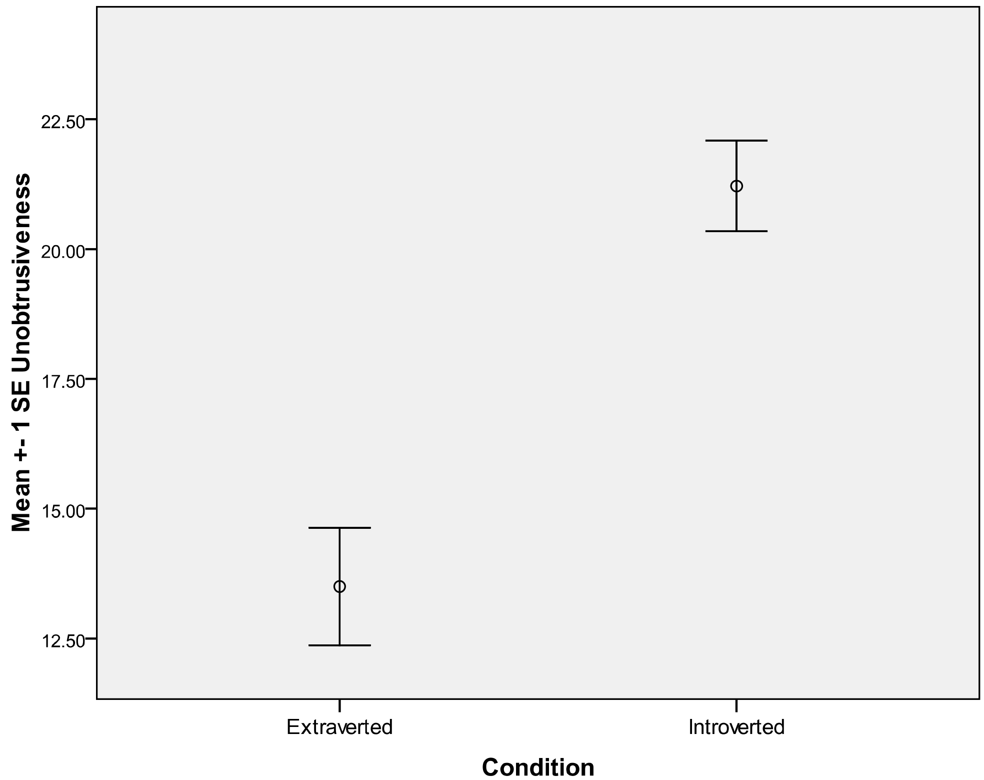


Figure 95: Standard Error/Means Plot for *Unobtrusiveness* scale: the participants found the *Introverted* robot less distracting.

The results on the Ease scale and its subscales were not as clear-cut: 1-tailed independent-samples T-tests were significant at 0.05 levels only for the *Ease* scale (t_{Ease} = -1.7, p<0.048), and *Easy* (t_{easy} = -1.9, p<0.033) and *Short* (t_{short} = -1.9, p<0.034) subscales; see **Table 35** for descriptive statistics.

Table 35: Descriptive statistics for *Ease* scale

	Condition	N	Mean	Std. Deviation	Std. Error Mean
Easy	Extraverted	14	4.4286	.75593	.20203
	Introverted	14	4.8571	.36314	.09705
Simple	Extraverted	14	4.6429	.63332	.16926
	Introverted	14	4.8571	.36314	.09705
Undemanding	Extraverted	14	4.5000	.75955	.20300
	Introverted	14	4.8571	.36314	.09705

Table 35 (continued)

Short	Extraverted	14	4.1429	1.02711	.27451
	Introverted	14	4.7143	.46881	.12529
Basic	Extraverted	14	4.6429	.63332	.16926
	Introverted	14	4.7143	.61125	.16336
Ease, Overall	Extraverted	14	22.3571	3.17701	.84909
	Introverted	14	24.0000	1.61722	.43222

The overall average *TLX* score presents a stronger result on perceived difficulty of the math problem: the 2-tailed independent samples T-test result was significant at 0.01 level, with $t_{TLXtotal} = 3.8$, $p<0.001$. See **Table 36** for descriptives for both the overall and the weighted subscale scores, and **Figure 96** for a standard error/means plot showing that the participants in the *Extraverted* condition found the problem more demanding.

Table 36: Descriptive statistics for *TLX* measure

	Condition	N	Mean	Std. Deviation	Std. Error Mean
Mental Demand	Extraverted	13	21.3077	16.57500	4.59708
	Introverted	14	12.5000	12.63542	3.37696
Physical Demand	Extraverted	13	1.0000	2.23607	.62017
	Introverted	14	1.5714	1.94992	.52114
Temporal Demand	Extraverted	13	34.1538	17.85519	4.95214
	Introverted	14	8.5000	8.99359	2.40364
Performance	Extraverted	13	19.1538	14.60506	4.05071
	Introverted	14	6.8571	7.49212	2.00235
Effort	Extraverted	13	24.0769	16.14756	4.47853
	Introverted	14	14.9286	16.26447	4.34686
Frustration	Extraverted	13	23.3077	26.40197	7.32259
	Introverted	14	3.7143	3.72989	.99686
TLX Total	Extraverted	13	7.6872	3.69488	1.02478
	Introverted	14	3.2190	2.40433	.64258

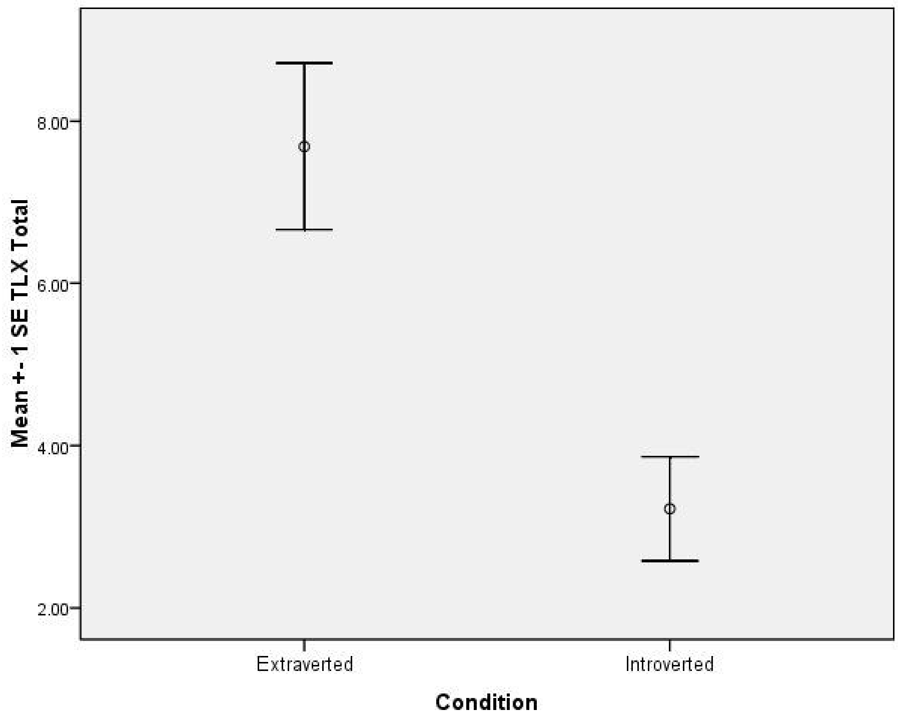


Figure 96: Standard Error/Means Plot for *TLX* score: the participants in the *Extraverted* condition found the math problem more demanding.

More interesting findings come from examining the subscales, as the distribution of ratings between them was different for each condition. In particular, there were no statistically significant differences in the ratings of *Mental Demand*, *Physical Demand*, and *Effort*, but the differences in the ratings of *Temporal Demand*, *Performance* and *Frustration* were significant at 0.01 and 0.05 levels. The relative ratings on the subscales are presented graphically in **Figure 97**: *Temporal Demand* was rated **4** times higher and *Frustration* **6.3** times higher when the *Extraverted* robot supervised the math problem. Based on these results our conjecture is that the participants may have felt **rushed** (*Temporal Demand*) and **frustrated** (*Frustration*) when a gregarious and energetic robot tried to engage them in small talk while they were concentrating on a problem solving task. Please note that the math task was found rather undemanding overall: the score of 7.69 out of 21 possible for *Extraverted* condition, and only 3.22 for *Introverted*. This could explain the weaker result on the *Ease* scale: it had only 5 points compared to 21 of *TLX*, therefore the difference, though still present, was less pronounced.

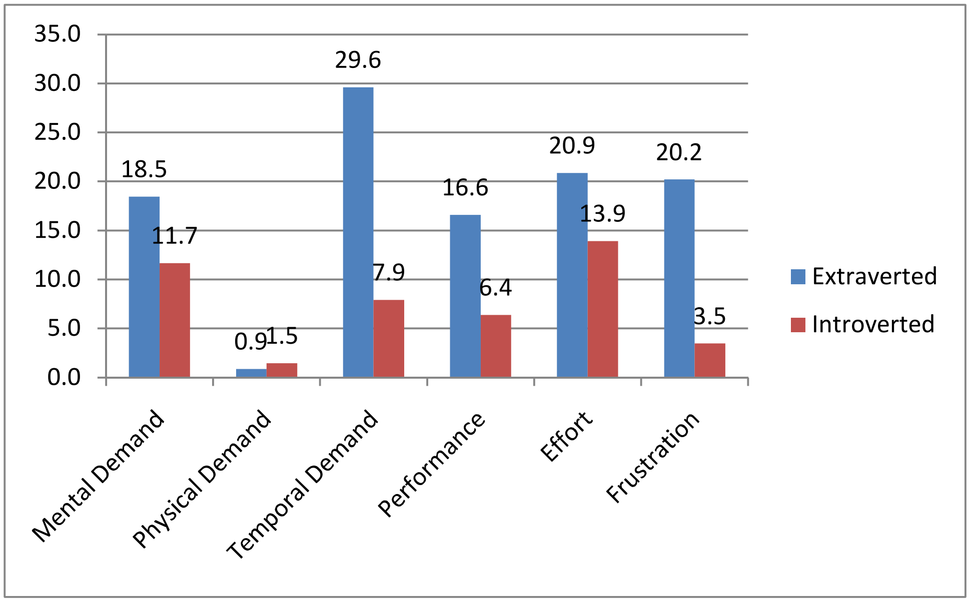


Figure 97: Distribution of weighted *TLX* subscale scores: Temporal Demand and Frustration were rated much higher in the *Extraverted* condition.

Additionally, Pearson's Correlation test was performed to compare *Appropriateness* (math task), *Unobtrusiveness*, *Ease* and *TLX* scales (**Table 37**). *TLX* and Ease ratings were strongly negatively correlated at the 0.01 level: the easier the subjects found the problem, the less demanding it appeared. *TLX* score was also strongly negatively correlated with both *Appropriateness* (math) and *Unobtrusiveness* scores (at the 0.01 level), suggesting that those who found the robot more appropriate and unobtrusive, also perceived the math problem as less demanding. Finally, *Unobtrusiveness* was strongly correlated with *Appropriateness* (at the 0.01 level) and moderately with *Ease* (at the 0.05 level).

Table 37: Correlations between *Appropriateness*, *Unobtrusiveness*, *Ease* and *TLX* scales

		Appropriateness (math)	Unobtrusiveness	Ease	TLX Total
Appropriateness (math)	Pearson Correlation	1	.807**	.370	-.685**
	Sig. (2-tailed)		.000	.052	.000
	N	28	28	28	27
Unobtrusiveness	Pearson Correlation	.807**	1	.392*	-.751**
	Sig. (2-tailed)	.000		.039	.000
	N	28	28	28	27
Ease	Pearson Correlation	.370	.392*	1	-.518**
	Sig. (2-tailed)	.052	.039		.006
	N	28	28	28	27
TLX Total	Pearson Correlation	-.685**	-.751**	-.518**	1
	Sig. (2-tailed)	.000	.000	.006	
	N	27	27	27	27

**. Correlation is significant at the 0.01 level (2-tailed).

*. Correlation is significant at the 0.05 level (2-tailed).

To summarize, the hypothesis stating that the *Introverted* robot would be found more appropriate for and unobtrusive during the math task, and that the task itself will be perceived as less demanding when proctored by the *Introverted* robot rather than its extraverted counterpart was confirmed. In particular,

- The *Introverted* robot was found to be more **appropriate**, **right for task**, **well-suited**, **proper** and **matched to task** than the *Extraverted* one with regards to the **math** task;
- The *Introverted* robot was also rated as **easier to tune out**, **quieter**, **less demanding and offensive**, and better at **minding its own business** than the *Extraverted* one;

- The math task was perceived as more **demanding** overall (based on TLX scores) when it was supervised by the Extraverted robot, and in particular it was rated higher on Temporal Demand, Performance and Frustration; also, those who perceived the robot as **more appropriate** for the math task and **unobtrusive** also found the problem itself **less demanding**.

6.2.5.4 Hypothesis 3: Task Performance

It was hypothesized that subjects' performance on the *quiz* task will be better in the Extraverted condition. This hypothesis was not confirmed, as either 2- or 1-tailed independent-samples T-tests on quiz score and quiz completion time did not show any significant difference between the conditions. No practice effects were observed on the quiz task – that is, neither the quiz score nor completion time improved significantly in the experimental task compared to the baseline task (see **Table 38** for descriptive statistics). Although the participants found the facts presented by the *Extraverted* robot more fun, exciting and interesting, this difference in perception did not result in objective improvement in *quiz* performance.

Table 38: Descriptive statistics for task performance

	Condition	N	Mean	Std. Deviation	Std. Error Mean
Quiz Score (experimental)	Extraverted	15	3.7333	.96115	.24817
	Introverted	15	4.0000	.84515	.21822
Quiz Score (baseline)	Extraverted	15	3.6000	1.12122	.28950
	Introverted	15	3.3333	.89974	.23231
Quiz Completion (experimental)	Extraverted	15	96.3333	45.12311	11.65074
	Introverted	15	74.4667	30.13272	7.78024
Quiz Completion (baseline)	Extraverted	15	76.6000	25.35970	6.54785
	Introverted	15	86.8667	25.51993	6.58922
Math Completion (baseline)	Extraverted	14	115.7143	36.69506	9.80717
	Introverted	14	138.3571	47.33938	12.65198

Table 38 (continued)

Math Completion (experimental)	Extraverted	14	89.2143	29.75070	7.95121
	Introverted	14	76.3571	22.60227	6.04071
Completion Time Improvement (math)	Extraverted	14	.2210	.17490	.04674
	Introverted	14	.4123	.22724	.06073

It was also hypothesized that subjects in *Introverted* condition would perform better on *math* task, as they would be less distracted by the robot. Overall, the participants performed exceedingly well on the math problem: only 2 out of 15 participants per condition solved the baseline problem incorrectly, and 1 out of 15 participants in *Extraverted* condition and 2 out of 15 in *Introverted* solved the experimental problem incorrectly; therefore, statistical analysis on correctness was not possible. However, some practice effects in terms of completion time were observed, as evidenced by a 2x2, Practice (Baseline vs. Experimental task) x Trait (Extraverted vs. Introverted) ANOVA: there was a significant main effect of Practice ($F_{Practice} = 22$, $p<0.000$) on completion time, suggesting that the experimental problem took less time to solve (**Figure 98**).

In order to examine this finding further, we calculated percentage of improvement in completion time, and found that Completion Time Improvement was more pronounced in the Introverted condition, as confirmed by a 2-tailed T-test, with $t_{CompletionTimeDifference} = -2.5$, $p<0.019$; see **Figure 99** for the standard error/means plot. This observation shows that the participants in the *Introverted* condition improved more from baseline to experimental task in terms of completion time, even though there was no statistically significant difference in completion time of experimental task per se. **Table 38** provides descriptive statistics for all performance metrics.

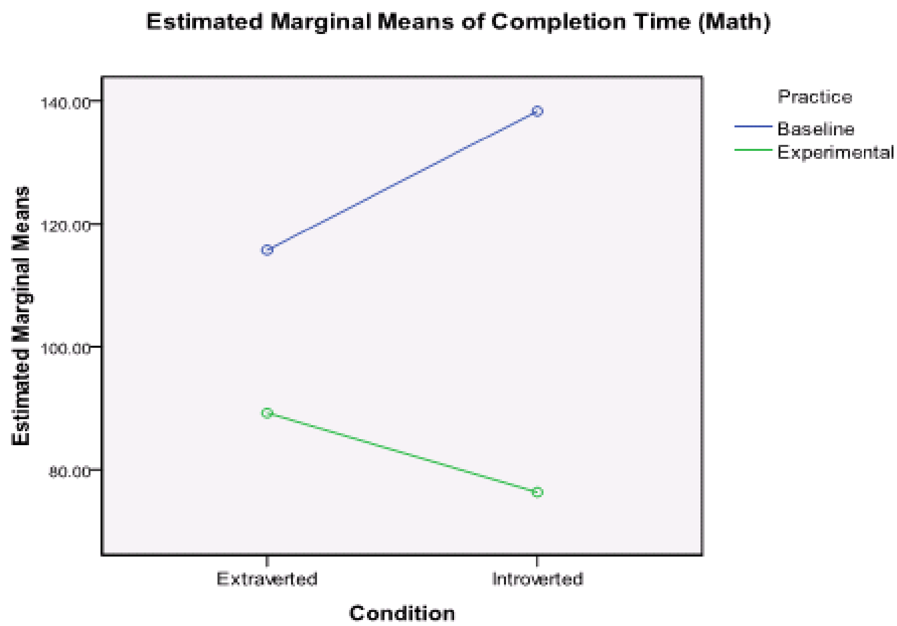


Figure 98: Practice x Trait means plot: regardless of the condition, it took participants less time to complete the experimental math task.

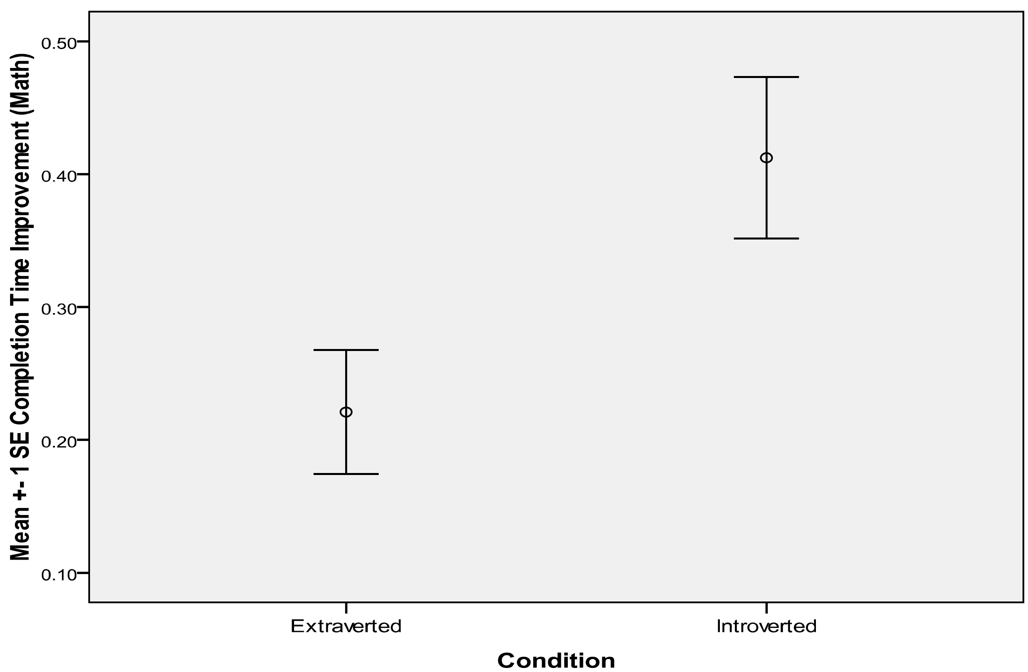


Figure 99: Standard Error/Means Plot for *Completion Time Improvement* for *math* task: the participants in the *Introverted* condition improved their completion time more.

Additionally, a significant main effect of Order (baseline vs. experimental) on Completion Time Improvement on the math task, as a result of 2x2 (Order x Trait)

ANOVA (F_{Order} = 4.8, p<0.039). It shows that the participants improved their time more when the math task was presented first (see **Table 39** for descriptive statistics, and **Figure 100** for a means plot).

Table 39: Descriptive statistics for Completion Time Improvement, Order x Trait

Condition	Order	Mean	Std. Deviation	N
Extraverted	Math First	.2902	.16781	8
	Quiz First	.1288	.14917	6
	Total	.2210	.17490	14
Introverted	Math First	.4917	.07705	7
	Quiz First	.3330	.30207	7
	Total	.4123	.22724	14
Total	Math First	.3842	.16570	15
	Quiz First	.2387	.25714	13
	Total	.3167	.22154	28

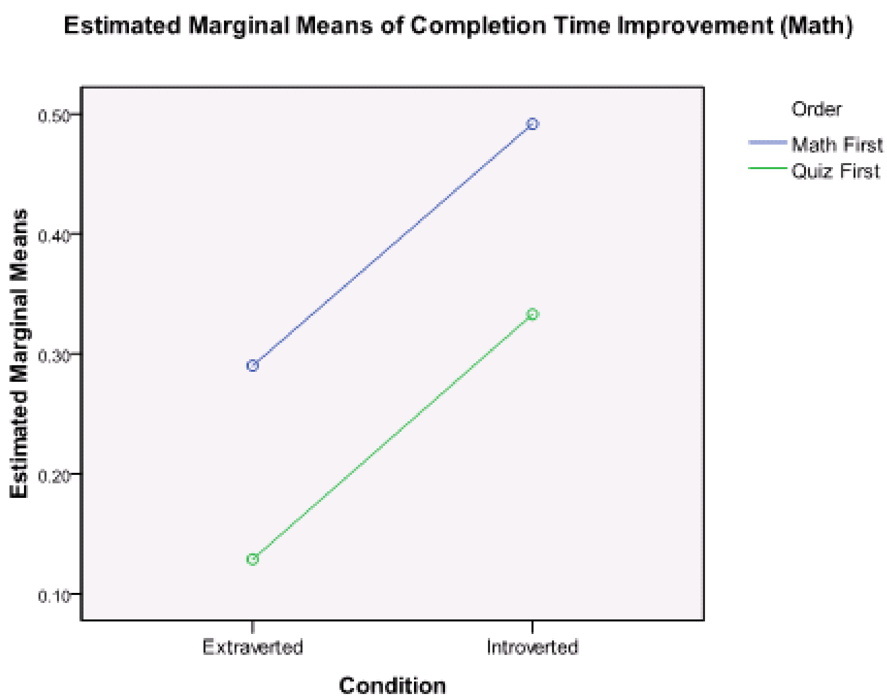


Figure 100: Order x Trait means plot for math Completion Time Improvement: the time improvement was less pronounced in both conditions when the quiz task was given first.

To summarize, the task performance hypothesis was confirmed for the math task, but not for the quiz. In particular:

- despite the *Extraverted* robot's better ratings as a presenter (the participants felt more welcome, and found the facts it presented more appealing), **no differences** were found on the **quiz performance**;

- for the *math* task, as predicted, the ***Extraverted*** robot's engaging and bold demeanor during the *math* task was detrimental, and resulted in a **lower** percentage of **completion time improvement** between the baseline and experimental tasks.

6.2.5.5 Hypothesis 4: Extraversion Recognition

It was hypothesized that the display of *Extraversion* and *Introversion* will be recognized on a humanoid robot. Indeed, this hypothesis was confirmed, as the participants rated the *Extraverted* robot as significantly higher on Extraversion scale of the robot personality questionnaire (Appendix FF) than the *Introverted* robot, as evidenced by the results of a 2-tailed T-test, with $t_{Extraversion} = 4.5$, $p<0.000$. In particular, they found Extraverted robot very high on Extraversion (mean score 7.38 out of 9, SD = 0.81), whereas the robot in the *Introverted* condition was rated as medium (mean score 5.74 out of 9, SD = 1.14); see **Figure 101** for a plot.

Pearson's Correlations test was also performed to determine whether the participants projected their personality on the robot, and no correlation was found between the subjects' and robot's ratings on Extraversion, suggesting that they rated the robot according to its own personality manifestation.

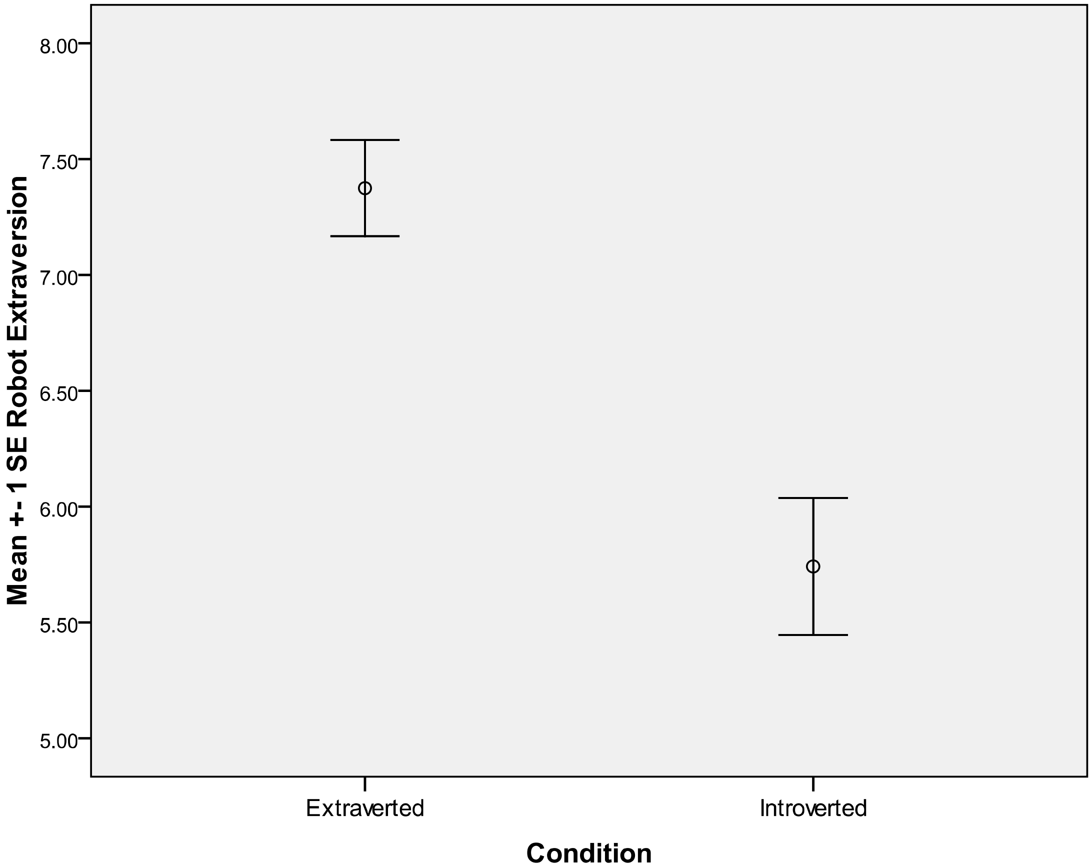


Figure 101: Standard Error/Means Plot for *Extraversion* score: the *Introverted* robot was rated significantly lower on the Extraversion dimension of the personality test.

6.2.5.6 Other observations

Although the following observations did not address the study hypotheses directly, they are nonetheless useful to obtain a more detailed picture of the results:

- No predictions were made a priori regarding the perception of *Naturalness*, as both *Introverted* and *Extroverted* individuals are common. No significant differences on this scale or its subscale were found, with one notable exception: the *Extraverted* robot was rated as more *Humanlike*, as evidenced by the result of a 2-tailed T-test, with $t_{humanlike} = 3.1$, $p<0.004$ (see **Table 40** for descriptive statistics).

Table 40: Descriptive statistics for *Naturalness* scale

	Condition	N	Mean	Std. Deviation	Std. Error Mean
Natural	Extraverted	15	3.3333	1.17514	.30342
	Introverted	15	2.9333	1.09978	.28396
Humanlike	Extraverted	15	2.9333	1.03280	.26667
	Introverted	15	1.9333	.70373	.18170
Conscious	Extraverted	14	3.2857	.82542	.22060
	Introverted	15	2.8000	1.08233	.27946
Lifelike	Extraverted	15	2.9333	.88372	.22817
	Introverted	15	2.4667	.83381	.21529
Animate	Extraverted	15	3.8667	.83381	.21529
	Introverted	15	3.8000	1.01419	.26186
Naturalness, Overall	Extraverted	15	16.1333	4.08598	1.05500
	Introverted	15	13.9333	3.30512	.85338

In addition, there was a positive correlation at the 0.05 level between the ratings of *Humanlike* and those of *Quiz Appropriateness*, *Appeal* and robot *Extraversion,* suggesting that the participants who found the robot more humanlike, also found it more appropriate for the *quiz* task, its presentation more appealing, and its extraversion more pronounced. **Table 41** presents these findings.

Table 41: Correlations between the *Humanlike* subscale, the scales of *Quiz Appropriateness* and *Appeal*, and *Extraversion* score

		Humanlike	Appropriate (quiz)	Appeal	Robot Extraversion
Humanlike	Pearson Correlation	1	.375*	.454*	.362*
	Sig. (2-tailed)		.041	.012	.050
	N	30	30	30	30
Appropriate (quiz)	Pearson Correlation	.375*	1	.441*	.292
	Sig. (2-tailed)	.041		.015	.117
	N	30	30	30	30

Table 41 (continued)

Appeal	Pearson Correlation	.454*	.441*	1	.387*
	Sig. (2-tailed)	.012	.015		.034
	N	30	30	30	30
Robot Extraversion	Pearson Correlation	.362*	.292	.387*	1
	Sig. (2-tailed)	.050	.117	.034	
	N	30	30	30	30

*. Correlation is significant at the 0.05 level (2-tailed).

- Pearson's correlations test revealed a strong negative correlation (at the 0.01 level) between *Extraversion* and *Completion Time Improvement* in the *math* task, and a moderate correlation (at the 0.05 level) between Extraversion and Completion Time for experimental math problem (**Table 42**). This indicates that those who perceived the robot as **more** *Extraverted*, improved **less** between the baseline and experimental math problems, and took **longer** to solve the problem.

Table 42: Correlations between *Extraversion*, math *Completion Time*, and *Improvement* in math *Completion Time*

		Math Completion (experimental)	Completion Time Improvement (Math)	Robot Extraversion
Math Completion (experimental)	Pearson Correlation	1	-.585**	.434*
	Sig. (2-tailed)		.001	.021
	N	28	28	28
Completion Time Improvement (Math)	Pearson Correlation	-.585**	1	-.601**
	Sig. (2-tailed)	.001		.001
	N	28	28	28
Robot Extraversion	Pearson Correlation	.434*	-.601**	1
	Sig. (2-tailed)	.021	.001	
	N	28	28	30

**. Correlation is significant at the 0.01 level (2-tailed); *. Correlation is significant at the 0.05 level (2-tailed).

To summarize, the following observations outside of the previously stated hypotheses were made:

- the *Extraverted* robot rated as more **humanlike** than the *Introverted* one;
- those participants who found the robot more **humanlike**, also found it more **appropriate** for the ***quiz*** task, its presentation more **appealing**, and its **extraversion more pronounced**;
- those who perceived the robot as **more** *Extraverted*, improved **less** between the baseline and experimental math problems, and took **longer** to solve the problem.

6.2.6 SUMMARY AND DISCUSSION

The goal of this experiment was to determine whether certain personality traits are better suitable for certain types of tasks. The first three hypotheses posited a priori for this study were with regards to this overall objective. The results of the experiment have validated all three, to a large extent. In particular:

- The ***Extraverted*** robot made the participants feel more **welcome** (overall) during the quiz task, and made the same building demolition facts appear more **appealing**. This capability would certainly be welcome for a robot engaging in intrinsically people-oriented jobs, such as a receptionist, a guide (museum or otherwise), a nurse bot.
- However, the same characteristics that were well-suited for an engaging task were detrimental to one requiring concentration. In particular, the ***Extraverted*** robot was found to be less appropriate and **more obtrusive** (overall) during the ***math*** task, and the problem it supervised appeared more demanding; it also affected the performance on the *math* task negatively:

Those were the expected results; the rest of the subsection will concentrate on the findings that were less straightforward.

1. **Quiz Appropriateness.** There was a strong Order effect on Appropriateness ratings for the quiz task. In case of *Introverted* robot, it was rated significantly less appropriate for the quiz task (as neither very appropriate nor very inappropriate) **only** when the quiz task followed the math. On the other hand, the task order didn't make much of a difference in the *Extraverted* condition: the robot was rated as equally highly appropriate. It is difficult to make a conjecture at this point of why this was the case, and additional experimentation with more participants would help uncover the underlying cause. However, it is possible that the *Introverted* robot's apparent suitability for the math task made its shortcomings during the presentation more obvious.

2. **Quiz Performance.** No differences were observed between the conditions on quiz performance. It appears that subjective perceptions of feeling more welcome and enjoying the presentation better in the *Extraverted* condition did not directly translate into objective performance improvement. There could be a number of reasons behind this finding:

   - The advantages of extraverted personality in teaching and leadership may be more pronounced over a longer span, in which case the short presentation simply was not enough to produce such benefits;

   - The *Extraverted* robot's frequent gestures and pose shifts may have been distracting to an extent, forcing the participants to pay attention to the robot's behavior rather than words. This reason was mentioned informally after the experiments by a number of the participants, and the disadvantage could possibly be ameliorated by more prolonged exposure, when the novelty effect wears off. On the contrary, the participants in the *Introverted* condition reported (informally, in conversations with the experimenter after the session was over) noticing their mind wander during the presentation, thus also causing them to pay less attention, which is unlikely to change with longer interaction.

6.3 SUMMARY

This chapter described design, implementation, administration and results of two HRI experiments intended to evaluate formally a subset of affective phenomena modeled in the *TAME* system. Both experiments identified potential benefits of including affect into robotic systems: Negative Mood and Fear in the first study and trait of Extraversion in the second.

In particular, the goal of the first experiment was to determine whether robotic expressions of mood and emotions – Negative Mood and Fear in this specific case – may provide identifiable benefits for human-robot interaction. The results of the study showed a number of advantages the affective robot had over its non-affective counterpart:

1. The participants found the robot's request to evacuate more **compelling**, **sincere** and **convincing** in one or both affective conditions than in control.
2. They complied with the robot's request to "evacuate" to a greater extent in the affective conditions:
    - the subjects were **faster** in complying with the robot's request to leave the "dangerous" zone (in the *Combined* condition);
    - they were **more prone** to respond to an **indirect request** to evacuate in both of the affective conditions;

- **more** of those in the affective conditions **walked further** towards the exit than in the control.

3. The participants reported feeling **more nervous after** interacting with the robot in the *Combined* condition than in *Control*, potentially making them more alert to any unfavorable changing in the surroundings.

The overarching goal of the second experiment was to determine whether certain personality traits are better suitable for certain types of tasks, and would thus be beneficial for human-robot interaction. The results showed that the *Extraverted* robot was found to be better suited for a people-oriented task – giving a presentation as part of an exhibit tour, and vice versa, the *Introverted* robot was more appropriate for a task requiring concentration from the participants - math task. These finding are summarized below:

1. *Extraverted* Robot was more suited for the *quiz* task:
    - The participants reported that the *Extraverted* robot in the *quiz* task made them feel as if their presence during the tour was more **liked**, **encouraged**, **wanted** and **desirable**.
    - The **facts** presented by the *Extraverted* robot appeared more **fun**, **exciting** and **interesting** than those given by the *Introverted* robot.
2. *Introverted* Robot was more suited for the *math* task:
    - The *Introverted* robot was found to be more **appropriate**, **right for task**, **well-suited**, **proper** and **matched to task** than the *Extraverted* one with regards to the **math** task.

- The *Introverted* robot was also rated as **easier to tune out**, **quieter**, **less demanding and offensive**, and better at **minding its own business** than the *Extraverted* one.
- The math task was perceived as less **demanding** overall (based on TLX scores) when it was supervised by the *Introverted* robot, and in particular it was rated **lower** on *Temporal Demand*, (concern over) *Performance* and *Frustration*; also, those who perceived the robot as **more appropriate** for the *math* task and **unobtrusive** also found the problem itself **less demanding.**
- *Introverted* robot's performance as a *math* task proctor resulted in a **higher** percentage of **completion time improvement** between the baseline and experimental tasks.

Finally, one final observation is worth emphasizing. Based on our experience in evaluating robot affect in the context of human-robot interaction, people may not always be able to recognize affective robotic expressions on a conscious, reportable level, especially from short interactions. However, active recognition may **not** be necessary to obtain a desired response; for example, although the participants in the mood and emotion experiment did not report any significant differences between the robot expressing negative mood and/or fear, and the one that was not, they still reacted to the cues sent out through the display of affect by complying with the robot's request to a greater extent and rating it as more compelling. This finding does not obviate the need to test the recognition of affective nonverbal robotic behaviors prior to conducting human-robot interaction evaluations; on the contrary, such testing is a good experimental practice, as it would be more likely to produce successful recognition results due to a greater salience the affective behavior would obtain in a more focused study.

This difficulty in robot affect recognition may stem in part from the subtle and volatile nature of affect: the relative infrequency and short duration of emotions, subtlety of moods, and slowly changing nature of attitudes. All these make evaluating affective robotic behavior through HRI studies challenging, though not impossible. Longer interaction time, exaggerated affective expressions, use of multiple affective phenomena (where applicable) and a priori robot affect recognition studies would help overcome this challenge.

7 MEASURING SOCIAL RESPONSE TO AFFECTIVE ROBOTS

One of the major challenges facing affective HRI is effective 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: for example, whether the presence of a robot as a partner improved the task completion time and by what percent, or whether a certain algorithm made learning faster or less error-prone. In the case of affective robots, however, it is not the robot performance per se that needs to be evaluated, but rather the social response the robots invoke in people they interact with. Do people find certain affective behaviors in robots more persuasive, natural and welcoming than others? Does robotic personality make some collaborative human-robot tasks seem more appealing and less arduous? Are these differences in subjective perception reflected in people's compliance with robot's requests and in task performance? The social response to affect in robots may be measured through a variety of means, both subjective and objective. This chapter will provide an overview of such HRI measures, and a more detailed discussion of the metrics used in the experiments described in Chapters 4 and 6, thus directly addressing the third research subquestion, "*What are the metrics for evaluating affective robotic behavior?"*

7.1 HRI MEASURES OVERVIEW

In their review of human study methods in HRI, Bethel et al. [65] point out five types of measures commonly used for evaluation in the human-robot interaction community: *self-assessments*, *interviews*, *observational (behavioral) measures*, *psychophysiology measurements*, and *task performance metrics*. In this subsection, each type will be discussed briefly, and its advantages and disadvantages will be highlighted.

7.1.1 SELF ASSESSMENTS

These are subjective metrics used to uncover people's perceptions of and attitudes towards their interactions with robots. These methods of evaluation commonly include: Likert-style questionnaires designed for evaluating specific goals of a particular study (often ad hoc in the HRI community [80]); 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 psychological and sociological measures, borrowed from corresponding research communities.

These latter tests have been developed and validated for use on human subjects, and can be employed to assess subjects' mood, emotional state, attitudes, presence, acceptance, and many other subjective states, but they have not as yet been adapted to and validated for use in the robotics domain. Examples of such measurement scales include: A Brief Version of Goldberg's Unipolar Big-Five Markers (personality) [163], Positive/Negative Affect Schedule (current mood state) [162], Self-Assessment Manikin (emotional response) [190], and the International Affective Picture System (emotional response) [191] . The aforementioned instruments are particularly suitable for evaluating affective robotics, as they can be used to both assess participants affective state, and test their the recognition of robot affect. The first two of these measurement tests have been used successfully in a number of *TAME* HRI experiments ([107, 165] and subsections 6.1, *Evaluating Expressions of Negative Mood and Fear in a Search-and-Rescue Scenario* and 6.2, *Evaluating Expressions of Extraversion and Introversion in a Robot as a Guide Scenario* of this dissertation).

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 in a number of ways:

- They are notoriously unreliable, as they might depend on a subject's cultural and educational background, age, gender, religious beliefs, current mood and motivational state, reasons for participating in the study, prior knowledge and attitudes, and a slew of other individual differences. In our experience, people's expectations of what the robot should (or should not) look like and be capable of colored their responses to a great degree: for example, if a person didn't think modern robots could talk or walk, and then observed a robot do exactly that, their opinion of the robot would be higher. Similarly, people differed greatly in how they interpreted the questions or scales – a person could give praises on how lifelike the robot was in their informal comments after the Search-and-Rescue experiment, and yet rate it as neither artificial nor lifelike, at best, in the questionnaire; and vice versa. Finally, self-assessments may suffer from acquiescence bias, where participants are more likely to agree than disagree with a statement, or might adjust their answers towards those with positive connotation [192].

- Replication of results and comparison between different studies is difficult; this is especially true of questionnaires put together in an ad hoc manner to suit a particular study. Such questionnaires are often too specific, and would not be readily applicable to other experiments, thus the results would not be comparable or repeatable. Reusable scales measuring concepts of common applicability to HRI would partially ameliorate this problem, and Bartneck et al. [80] advocate the use of such scales over Likert-style questions in developing questionnaires. Finally, although accepted psychological measurement instruments do have the advantage of confirmed validity across a large number of participants, they may not necessarily be valid for HRI, and at least some tests were found to be less reliable when used to rate robots versus humans [68].

7.1.2 INTERVIEWS

This is another subjective method for obtaining user perceptions, opinions and attitudes. The advantage of an interview over a questionnaire is the depth and breadth of exploration, especially with semi- or unstructured interviews: for example, where a questionnaire might ask a respondent to rate a robot's persuasiveness, an interview could go further to uncover what in the robot's appearance, behavior or words made it more or less persuasive, and what the interviewee means by "persuasiveness". Unfortunately, although interviews may provide a bigger picture and a more detailed account, they are even harder to compare between studies, or even between participants in the same study. They are also more time- and resource-consuming than self-assessments, as they require at least 2 trained independent raters to obtain reliable data evaluation. However, even without the formal rating, they can provide information to the experimenter regarding any confusion participants may have had about the study, or any unusual views they might hold.

7.1.3 BEHAVIORAL MEASURES

These 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 behaviors to watch for are carefully selected and accurately described, and then are extracted from the video either automatically, or by independent human coders. For example, suppose that the duration of mutual gaze is a good predictor of the quality of interaction – the longer the mutual gaze episodes, the more pleasant the interaction. Now we have a quantitative measure that would allow us to compare between robots that express affect and those that don't.

As this method does not rely on self-report, it is not subject to the same user bias as self-assessments and interviews; however, individual differences in behavioral styles will still make it difficult to compare across the participants. From our experience with this method in the AIBO study, some participants were extremely talkative and engaging in their interactions with the robotic dog, while others hardly uttered one or two phrases, and did not participate beyond the required instructions. Unless the number of people taking part in a study is rather large, these individual differences may be greater than those imposed by experimental conditions.

This method also suffers from interpretation bias: the definition of what a particular behavior or expression consists of (e.g., angles, acceptable percent of deviation, minimum duration, etc. for a mutual gaze) needs to be determined ahead of time and adapted to the current experiment, as well as what these macro or micro behaviors mean in respect to study hypotheses. Finally, another disadvantage of this method is that the expense of having independent coders analyze the video recordings becomes prohibitive for many research groups.

7.1.4 PSYCHOPHYSIOLOGY MEASURES

In this method, certain physiological responses (such as heart rate, skin conductance and temperature) can be measured before, during and after the interaction; such responses can be correlated with subjects' emotional state and arousal level. 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 biggest disadvantage of this method is the limitations 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 the level of anxiety needs to be determined [193, 194], but would need to be supplemented by other measures to obtain cross-validation and additional information. In addition, psychophysiology measures will suffer from individual differences in autonomic responses, as well as low reliability unless the equipment is individually calibrated. Finally, the equipment is often cumbersome and its presence alone may influence the results.

7.1.5 TASK PERFORMANCE METRICS

Objective task-related measures allow quantifying benefits a particular robot type/behavior/algorithm might have through such variables as accuracy, performance success, task completion time, error rate, resource usage and others, depending on a particular task and scenario. A clear-cut advantage of this method is the removal, to a large extent, of both subject and interpretation bias. From the point of view of affective HRI, it means measuring changes in human task performance occurring as a response to changes in the robot introduced by experimental conditions, rather than robot performance directly.

One notable example of employing a task performance metric to evaluate the effectiveness of robot affect is presented in Scheutz et al. [195] . In this 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 (expressed speech rate and pitch) increased, the participants were alerted to the impending deadline, and worked more efficiently. Compliance is another task performance metric that has been previously employed in HIR. It was used to evaluate the effect of robotic personality (playful or serious) on participants' compliance with the robot's request to create and perform an exercise routine, measured by the amount of time participants exercised by themselves [196].

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 them.

7.1.6 HRI MEASURES SUMMARY

To summarize, each type of HRI measures described in this subsection has its own set of advantages and disadvantages. Self-assessments and interviews provide a wealth of information about participant's perceptions of and attitudes towards robots, but suffer greatly from subjectivity and user bias. Behavioral analysis provides a different view of interaction and partially reduces subject bias, but is time- and resource-consuming, and is prone to interpretation bias. Psychophysiology methods provide a good indication of users' arousal in real time, but are costly and cumbersome to set up, and the information they produce is rather limited. Finally, task performance metrics are objective, but reflect only one side of the story – how well the participants could perform a task, rather than how satisfactory, easy or pleasant their interaction with a robot was. Bethel et al. [65] advocate that no single measurement is sufficient to evaluate any interaction, and that it is important to include more than one method of evaluation to obtain comprehensive understanding and convergent validity.

7.2 HRI METRICS EMPLOYED IN TAME EXPERIMENTS

In the course of this dissertation, three HRI studies and one online survey have been performed. The measures used to evaluate robot affect in these experiments ranged from questionnaires (Likert-style and open-ended questions and semantic differential scales) to established psychometric tools to task performance metrics. The rest of this subsection will focus on the metrics specifically suitable for affective HRI and which present a contribution to the field.

7.2.1 ESTABLISHED PSYCHOMETRIC MEASUREMENT TOOLS

Two psychological tools for measuring mood (PANAS [162]) and personality ("Big-Five Mini-Markers" [163]) were employed with moderate success in multiple studies, to assess both subjects' affect, and their recognition (or attribution) of robot affect. These are self-assessment tools with proven reliability and validity [162, 163, 197] , but which have not been systematically used in the HRI research. Both are discussed in more detail below.

7.2.1.1 PANAS: Positive Affect / Negative Affect Schedule

PANAS scales were developed by Watson et al. [162] to measure two aspects of mood: Positive Affect and Negative Affect, and are applicable for measuring mood for different time frames, from "right now" to "today" to "past few weeks" to "in general". The tool consists of two separately scored (but intermixed when presented to respondents) 10-item scales corresponding to two mood dimensions. Positive Affect (PA) scale consists of the following adjective subscales: *interested, excited, strong, enthusiastic, proud, alert, inspired, determined, attentive, active*; Negative Affect (NA) scale is comprised of the following items: *distressed, upset, guilty, scared, hostile, irritable, ashamed, nervous, jittery, afraid*. For the studies in this dissertation, a variation of PANAS tool, PANAS-T, was used, which contained 2 additional items, one for each scale: *happy* for PA, and *depressed* for NA. For each 5-point subscale the subjects were asked to rate the extent to which they experienced each mood state during the specified time frame, ranging from "*very slightly or not at all"* to "*extremely"*. For this dissertation, "right now" time frame was used, as our concern was subjects' or robot's current affective state; see Appendix G for the mood questionnaire as applied to a person, and Appendix S for one with regards to a robot.

PANAS-T was used to assess participants' affect in the AIBO and Search-and-Rescue experiments (Chapter 4, *Exploratory Experimental Study*, and subsection 6.1, *Evaluating Expressions of Negative Mood and Fear in a Search-and-Rescue Scenario*, respectively) and to assess robot affect in the Search-and-Rescue experiment and robot affect recognition survey (subsection 5.3, *Online Survey on Recognition of Affective Robotic Behavior; a shortened version of the test was used*). In the longitudinal AIBO study, the subjects completed PANAS-T at the end of each of 4 sessions; it was found that lower Negative Affect was reported on average in the emotional condition (in which the robotic dog was programmed to display emotions of Joy, Interest, Fear and Anger, and a generally extraverted and agreeable personality). In the recognition survey, the respondents could differentiate between positive and negative mood expressed by a humanoid robot in two video clips. Finally, in the Search-and-Rescue experiment, the participants reported higher levels of Negative Affect and nervousness in the Combined (negative mood + fear) condition. These results suggest that PANAS can be successfully used for measuring mood state in HRI studies. However, we also observed some interesting phenomena, which need to be taken into consideration while using this tool in the future.

1. **Individual Subscales**. It should be noted that PA and NA scales of the PANAS measurement tool assess various aspects of positive and negative mood. For example, NA scale has hostile, ashamed and irritable subscales which may not be relevant to an experiment objective (e.g., induction of anxiety). In such a case, it would be worthwhile to examine individual subscales which would be more in line with the nature of the study. In particular, there was a greater difference between conditions on the *nervous* subscale of the NA scale, than on the overall scale in the Search-and-Rescue experiment (see subsection 6.1.5.4, *Hypothesis 3: Participants' Negative Affect* for details).

Similarly, the PA scale contains both valence and activation terms, with the latter reflecting engagement with the stimulus, but not evaluation of it. For example, Patrick et al. [198] found an increase in Positive Affect when they showed their participants negative pictures, which was due to the activation items of the PA scale. In the Search-and-Rescue experiment, we observed that the participants viewed the robot as more **determined** and **active** in one or both of the affective conditions. In particular, ANOVAs on the *determined* and *active* subscales were significant ($f_{determined} = 3.64$, $p < 0.035$, $f_{active} = 6.12$, $p < 0.005$), and LSD post-hoc comparisons revealed that the robot was rated as more determined in the *Mood* ($p < 0.022$) and *Combined* ($p < 0.027$) conditions than *Control*, and more active in the *Combined* condition that in either Mood ($p<0.021$) or *Control* ($p<0.008$). **Table 43** provides descriptive statistics for this finding.

Table 43: Descriptive Statistics for Robot *determined* and *active* subscales of the PA scale

Dependent Variable	Condition	N	Mean	Std. Deviation	Std. Error
Robot Active	Control	14	2.9286	.82874	.22149
	Mood Only	14	3.0714	1.07161	.28640
	Mood and Emotion	15	4.0667	.96115	.24817
	Total	43	3.3721	1.06956	.16311
Robot Determined	Control	14	2.7857	1.31140	.35049
	Mood Only	14	3.8571	1.02711	.27451
	Mood and Emotion	15	3.8000	1.20712	.31168
	Total	43	3.4884	1.26061	.19224

As a result, our recommendation with respect to individual subscales of the PANAS tool would be to: 1) be as specific as possible while outlining hypotheses regarding positive and negative affect; 2) pay attention to individual subscales in addition to overall PA and NA scores.

Projection of Affective State. During the analysis of the Search-and-Rescue experimental data on participant's and robot mood, it was noted that the subjects may have been projecting their positive affective state on the robot. In particular, there were strong positive correlations at the 0.01 level between the participants' ratings of their own positive affect **both before and after** the interaction, and the robot's positive affect (**Table 44**).

Table 44: Pearson's Correlations between participants' and robot's positive affect ratings: strong correlations between subjects' positive affect ratings before and after the interaction and the robot's positive affect ratings indicate projection of subject affect on to the robot

		Subject Positive Affect, After	Subject Positive Affect, Before	Robot Positive Affect
Subject Positive Affect, After	Pearson Correlation	1	.753**	.678**
	Sig. (2-tailed)		.000	.000
	N	43	43	43
Subject Positive Affect, Before	Pearson Correlation	.753**	1	.510**
	Sig. (2-tailed)	.000		.000
	N	43	43	43
Robot Positive Affect	Pearson Correlation	.678**	.510**	1
	Sig. (2-tailed)	.000	.000	
	N	43	43	43

**. Correlation is significant at the 0.01 level (2-tailed).

Interestingly, the same phenomenon was not observed in case of Negative Affect: there were significant correlations between the ratings of participants' and robot's negative mood **after** the interaction, but not **before** (**Table 45**). This suggests that in this particular experiment the negative affect may have been induced on the participants through the interaction, as opposed to being projected from people onto the robot.

Table 45: Pearson's Correlations between participants' and robot's negative affect ratings

		Subject Nervous, After	Subject Nervous, Before	Subject Negative Affect, After	Subject Negative Affect, Before	Robot Nervous	Robot Negative Affect
Subject Nervous, After	Pearson Correlation	1	.328*	.830**	.386*	.389*	.348*
	Sig. (2-tailed)		.041	.000	.012	.011	.024
	N	42	39	42	42	42	42
Subject Nervous, Before	Pearson Correlation	.328*	1	.219	.595**	-.024	-.075
	Sig. (2-tailed)	.041		.181	.000	.881	.648
	N	39	40	39	40	40	40
Subject Negative Affect, After	Pearson Correlation	.830**	.219	1	.393*	.267	.366*
	Sig. (2-tailed)	.000	.181		.010	.087	.017
	N	42	39	42	42	42	42
Subject Negative Affect, Before	Pearson Correlation	.386*	.595**	.393*	1	.121	.072
	Sig. (2-tailed)	.012	.000	.010		.441	.645
	N	42	40	42	43	43	43
Robot Nervous	Pearson Correlation	.389*	-.024	.267	.121	1	.715**
	Sig. (2-tailed)	.011	.881	.087	.441		.000
	N	42	40	42	43	43	43
Robot Negative Affect	Pearson Correlation	.348*	-.075	.366*	.072	.715**	1
	Sig. (2-tailed)	.024	.648	.017	.645	.000	
	N	42	40	42	43	43	43

*. Correlation is significant at the 0.05 level (2-tailed).

**. Correlation is significant at the 0.01 level (2-tailed).

We have noted a similar tendency in the AIBO study, where the participants reported AIBO as having emotions when none were expressly exhibited by the robot. Together, these findings point to a tendency of people to **project their**

own affective state on a robot **if** the robot does not exhibit any specific affect otherwise.

7.2.1.2 Big-Five Mini-Markers: a Brief Measure of Personality

This measure is a brief version of Goldberg's Unipolar Big-Five Markers [199] developed by Saucier [163] to reduce the burden on the respondents without sacrificing the robustness. The Mini-Markers consist of 40 adjectives, 8 per each dimension of the Five-Factor Model of personality, and respondents are asked to rate themselves or others on a 9-point scale for each attribute.

Mini-Markers were used to measure both participants' and robot's personality in two experiments: the AIBO study and the Extraversion experiment, and the Extraversion subset of the tool was used for determining recognition of a humanoid robot's level of extraversion in the robot affect recognition online survey (subsection 5.3, *Online Survey on Recognition of Affective Robotic Behavior*). Unlike with mood ratings, no correlations were observed between the ratings of participants' and robot's personality on corresponding dimensions, therefore suggesting that people are not likely to project their own personality on robots they interact with. On a lighter note, regardless of subjects' personality, Nao robot was described as highly or moderately extraverted (depending on the condition), rather conscientious, and not particularly bright (though still more intelligent/open than AIBO).

To evaluate the appropriateness of using Mini-Markers with regards to robots, the internal consistency reliability (as measured by Cronbach's Alpha) of each dimension on ratings of robot's personality (as measured across AIBO and Extraversion experiments) was computed. It was found to be somewhat lower than what was reported for human data ([163], Cronbach's alpha ranging from 0.76 to 0.86). However, it was still acceptable (above 0.7, as recommended by Nunnally [200] and Bartneck et al. [80]) for dimensions of *Conscientiousness*, *Neuroticism* and *Openness*, and just below for dimensions of *Extraversion* and *Agreeableness* (**Table 46**). This decrease in internal reliability is not entirely unexpected, as this measure was applied to non-human entities rather than humans, and the number of cases was low (51). We also do not believe it low enough to preclude this tool from use in HRI experiments – on the contrary, the results are rather encouraging and we would recommend Mini-Markers for further use in evaluation of affective robots.

Table 46: Internal Consistency Reliability for Mini-Markers (robot's personality)

Dimension	Experiment	AIBO	Extraversion	Average
Extraversion (8 items)	Cronbach's Alpha	.606	.762	.684
	N	21	30	51
Agreeableness (8 items)	Cronbach's Alpha	.592	.657	.625
	N	21	30	51
Conscientiousness (8 items)	Cronbach's Alpha	.756	.655	.706
	N	21	30	51
Neuroticism (8 items)	Cronbach's Alpha	.716	.718	.717
	N	21	30	51
Intellect/Openness (8 items)	Cronbach's Alpha	.935	.768	.852
	N	21	30	51

7.2.1.3 Summary

In this subsection, the suitability of two established psychometric tools for measuring affect (PANAS for Positive and Negative Affect [162], and Mini-Markers for personality [163]) for use in HRI studies was examined. Based on our prior experience with these tests in a number of *TAME* HRI studies we can recommend them as adequate means for evaluating both subjects' and a robot's affective state and personality. In particular, the following observations and recommendations can be made regarding PANAS and Mini-Markers use in HRI experiments:

- Subscales of the PANAS tool measure different aspects of Positive and Negative Affect; for example, PA scale contains both valence and activation terms, with the latter reflecting engagement with the stimulus, but not evaluation of it [198]. Therefore analyzing individual subscales in addition to the overall ratings would provide finer-grained results, especially in cases where a particular aspect of mood is of primary interest (e.g., nervousness, hostility, excitement, etc.).
- People are prone to project their own affective state onto the robot they are interacting with, especially if the robot does not display any discernable affect otherwise. Interestingly, this phenomenon was not observed with personality traits. This finding should be taken into account when analyzing the results of these tests, in order to differentiate the recognition of affective behaviors from people's projection of their own affective state.
- Internal consistency reliability on Mini-Markers used to asses a robot's personality was comparable, if somewhat lower than when that reported for its use with human subjects [163], confirming the suitability of using this tool in HRI studies.

7.2.2 SEMANTIC DIFFERENTIAL SCALES

The semantic differential scale, devised originally by Osgood et al. [201] , is a self-assessment (rating) tool, and has been used frequently for measuring social attitudes and perceptions. Bartneck et al. [80] advocate its use for HRI evaluation over Likert style scales due to consistency of presentation and reduction of acquiescence bias (common to Likert style scales, which force a respondent to either agree/disagree with or report their like/dislike of a statement). In addition, once developed, these scales can be reused in other studies, thus allowing inter-study comparison.

Typically, semantic differential scale is a 5 to 9 point bipolar rating (sub)scale, with opposites at each end, and respondents are required to select the point that most closely reflects their opinion; this provides both extreme options as well as more neutral ones. By combining 3 to 10 (or sometimes even more) such subscales together, a composite scale expressing an overarching concept can be designed.

As with any evaluation measure, there are certain considerations that need to be taken into account in the development of semantic differential scales; in the design of the scales for *TAME* experiments, we paid attention to the points brought up by Al-Hindawe [202]. In particular, the following design decisions were made:

- Both complementary opposites (e.g., sincere – insincere, conscious – unconscious) and more subtle, gradable antonyms (e.g., entertaining – boring, distracting – easy to tune out) were used, as deemed appropriate. Complementary opposites are not always available, and simple negation may project an unintended meaning; for example, using a direct opposite of quiet, “loud” would not quite relate the idea of “distracting” as opposed to simply “loud”.
- 5 items (adjective pairs) per scale were chosen to provide enough information about the chosen concepts, yet not be overly tedious for the subjects to go through.
- In all the scales, negatively valenced adjectives were placed on the left, and positively – on the right. This was done for consistency, to reduce any errors due to unexpected (from the subjects’ point of view) reversal of polarity.
- Five-point scales (as opposed to 7- or 9-point), although course-grained, were chosen to reduce the burden on the respondents and make grading less tedious.

The rest of the subsection will discuss each scale and present internal reliability results. It should be noted, that although these scales were developed with robots in mind, they could be applicable to a wider domain, e.g., virtual or other embodied agents.

7.2.2.1 Semantic Differential Scales Used in Search-and-Rescue Experiment

Three 5-item semantic differential scales were employed in this experiment: *Understandability*, *Persuasiveness* and *Naturalness*. *Understandability* scale measures the extent to which a robot is perceived as understandable, and can refer to: robot's behavior, actions, speech, expressions, "state of mind", intentions, and other attributes. *Persuasiveness* scale measures to what extent a robot was found to be persuasive, and can be applied to: robot's request, message, speech, actions, etc. Finally, the *Naturalness* scale measures to what extent a robot is judged as natural, and can refer to either a robot as a whole, or its appearance, speech or behavior separately. This scale was not developed from scratch, but rather combines a number of subscales of two overlapping scales, *Anthropomorphism* and *Animacy*, presented in Bartneck et al. [80], and eliminates redundancy. **Table 47** shows the adjectival opposites for the scales, and the scales themselves, as presented to respondents, can be found in Appendix T.

Table 47: Adjectival Pairs comprising Understandability, Persuasiveness, and Naturalness Scales

Understandability	**Persuasiveness**	**Naturalness**
Confusing – Clear	Ignorable – Compelling	Fake – Natural
Unreadable – Easy to Read	Inappropriate – Appropriate	Machinelike – Humanlike
Inconsistent – Consistent	Ineffective – Persuasive	Unconscious – Conscious
Hard to Understand – Easy to Understand	Insincere – Sincere	Artificial – Lifelike
Inexpressive – Expressive	Unconvincing – Convincing	Inert – Interactive

Two types of statistical analysis were performed to evaluate these scales: 1) factor analysis (principal components) to determine whether all the subscales within a scale refer to the same construct; and 2) internal consistency reliability test (measured by Cronbach's Alpha) which reflects the homogeneity of the scale. As a result of factor analysis, two factors (dimensions) were extracted for both *Understandability* and *Naturalness* scales, and one for *Persuasiveness*. Further intra-scale correlations analysis showed that the "expressive – inexpressive" pair did not correlate with any other subscales within the *Understandability* scale, and removing this item resulted in a single dimension returned by a subsequent factor analysis. Similarly, removal of the "interactive – inert" pair (which was correlated with only one other subscale) from the *Naturalness* scale resulted in a single dimension, based on a subsequent factor analysis. This adjectival pair was replaced in the *Extraversion* experiment.

To determine internal consistency reliability, Cronbach's Alpha was computed for each scale, both for each experimental condition and the experiment overall (see **Table 48** for internal consistency results, and Appendix II for descriptive statistics).

Table 48: Internal Consistency Reliability for *Understandability*, *Persuasiveness* and *Naturalness* scales, by condition and overall. Overall, all scales had acceptable reliability.

Scale	Condition	Control	Mood	Combined	Overall
Understandability (5 items)	Cronbach's Alpha	.625	.810	.450	.654
	N	14	14	15	43
Persuasiveness (5 items)	Cronbach's Alpha	.825	.408	.830	.799
	N	14	14	.15	43
Naturalness (5 items)	Cronbach's Alpha	.828	.824	.632	.779
	N	14	14	15	43
Understandability (expressive excluded, 4 items)	Cronbach's Alpha	.716	.880	.254	.714
	N	14	14	15	43

Overall, the scales have acceptable internal consistency (after the “inexpressive – expressive” pair was removed from *Understandability* due to its poor intra-scale correlations rating), with Cronbach’s Alpha values ranging from 0.714 for the 4-item *Understandability* scale to 0.799 for *Persuasiveness*. Although in some conditions the reliability was lower, it could be due a small number of respondents (14 or 15 per condition), given that the overall results reflecting a larger number of participant are better.

To summarize, the novel semantic differential scales of *Understandability* (with the “expressive –inexpressive” pair removed), *Persuasiveness* and *Naturalness* were found to have acceptable internal consistency reliability (in an HRI experiment involving 45 participants), reflect concepts important for the HRI domain, and are flexible enough to be used in a variety of scenarios. It should be noted that the exact wording of the questions introducing each scale (e.g., “In your opinion, the robot’s request to leave was:” preceding the *Persuasiveness* scale) will need to be adjusted based on the experimental procedure and the hypotheses.

7.2.2.2 Semantic Differential Scales Used in Extraversion Experiment

Five 5-item semantic differential scales were employed in this experiment: *Appropriateness* (used for two different tasks, *quiz* and *math*), *Welcome, Appeal, Unobtrusiveness, Ease* and *Naturalness*. The *Appropriateness* scale measures the extent to which a robot is perceived as appropriate for a particular type of task or process; it can be used in regards to a robot as a whole, or its appearance, behavior, capabilities, and other attributes. *Welcome* scale measures to what extent a robot made participants feel welcome, and can be applied to, for example, participants’ participation in a joint task, their presence, offer of assistance, etc. The *Appeal* scale measures the extent to which participants find an activity involving a robot appealing; it can refer to facts or a presentation given by a robot, a meeting, a joint task, etc. To measure the extent to which a robot is perceived as distracting during a task, a meeting, or any other activity, a scale of unobtrusiveness was developed; the lower the score, the higher the distraction due to the robot, as the negatively valenced adjectives are anchored at "1". The *Ease* scale measures the perceived ease of a task, a problem, or a joint project. Finally, the same *Naturalness* scale was used as before, but due to the poor intra-scale correlation result, the "inert – interactive" pair was replaced with a different activity-related pair, "inanimate – animate". **Table 49** shows the adjectival opposites for the scales. The scales themselves, as presented to respondents, can be found in Appendices CC and DD, and the descriptive statistics for them in Appendix II.

Table 49: Adjectival Pairs Comprising *Appropriateness*, *Welcome*, *Appeal*, *Unobtrusiveness*, and *Ease* Scales

Appropriateness	Welcome	Appeal	Unobtrusiveness	Ease
Inappropriate – Appropriate	Unwelcome – Welcome	Boring – Interesting	Distracting – Easy to Tune Out	Hard – Easy
Wrong for Task – Right for Task	Undesired – Desirable	Not Fun – A lot of Fun	Interfering – Minding its Own Business	Complicated – Simple
Ill- Suited – Well-Suited	Disliked – Liked	Useless – Useful	Annoying – Inoffensive	Demanding – Undemanding
Improper – Proper	Tolerated – Encouraged	Dull – Exciting	Irritating – Undemanding	Long – Short
Mismatched – Matched to Task	Unwanted – Wanted	Tedious – Entertaining	Bothersome – Quiet	Complex – Basic

Similar to the Search-and-Rescue experiment scales, the same two types of statistical analysis were performed to evaluate the scales used in the Extraversion study. To identify whether any scales should be reduced further, Factor Analysis (principal component) was performed; each scale was found to be comprised of a single factor, reflecting the same concept. In order to determine the internal consistency reliability, Cronbach's Alpha was computed for each scale, both for each experimental condition and the experiment overall (**Table 50**). Overall, the alpha values showed moderate to high internal consistency for all scales, and the results per condition were all at the acceptable level as well. The internal consistency of the *Naturalness* scale was improved from 0.779 to 0.827 with replacement of the "interactive" item with "animate"; and only one factor was extracted by factor analysis for the modified scale, indicating that it reflects the measured construct better than the original one.

Table 50: Internal Consistency Reliability for *Appropriateness*, *Welcome*, *Appeal*, *Unobtrusiveness*, *Ease* and *Naturalness* scales, by condition and overall

Scale	Condition	Introverted	Extraverted	Overall
Quiz Appropriateness (5 items)	Cronbach's Alpha	.923	.902	.918
	N	15	15	30
Math Appropriateness (5 items)	Cronbach's Alpha	.885	.970	.966
	N	14	14	28
Welcome (5 items)	Cronbach's Alpha	.881	.896	.914
	N	15	15	30
Appeal (5 items)	Cronbach's Alpha	.796	.837	.848
	N	15	15	30
Unobtrusiveness (5 items)	Cronbach's Alpha	.847	.970	.927
	N	14	14	28
Ease (5 items)	Cronbach's Alpha	.777	..878	.865
	N	14	14	28
Naturalness (5 items)	Cronbach's Alpha	.724	.813	.827
	N	15	14	29

To summarize, the novel semantic differential scales of *Appropriateness*, *Welcome, Appeal, Unobtrusiveness, Ease* and *Naturalness* (where "interactive" item was replaced with "animate") were found to have high internal consistency reliability (in an HRI experiment involving 30 participants), reflect concepts important for the HRI domain, and are flexible enough to be used in a variety of scenarios.

7.2.3 TASK PERFORMANCE METRICS

Whenever both objective and subjective measures are used in a study, it is important to address the question of how these measures relate to each other, and whether differences in objective performance are reflected in subjective perceptions. In order to answer this question, we examined the correlations between task performance results and subjective scale ratings for the Search-and-Rescue and Extraversion experiments.

In the Search-and-Rescue experiment, the *Time To Cross* variable measured request compliance; specifically, how much time it took the participants from the time the robot issued the direct request "Please proceed to the exit" and their reaching the first cross marker (see subsection 6.1.5.3, *Hypotheses 2: Request Compliance* for details). This metric was found to be negatively, or inversely, correlated (at the 0.05 level, two-tailed) with only two subjective variables, both related to how persuasive the robot's request was found: *compelling* subscale of the *Persuasiveness* scale, and Decision to Leave variable (the extent to which robot's behavior influenced the participants' decision to leave); see **Table 51** for Pearson correlations results. This finding indicates that those who found the robot's request more compelling and the robot's behavior more of a factor in their decision to leave, were also faster to "evacuate".

Table 51: Correlations between *Time To Cross*, *Decision To Leave* and *Compelling* variables: those who found the robot more compelling and its behavior more influential in their decision to leave, were also faster to "evacuate".

		Time To Cross	Decision To Leave	Compelling
Time To Cross	Pearson Correlation	1	-.380*	-.407*
	Sig. (2-tailed)		.027	.017
	N	34	34	34
Decision To Leave	Pearson Correlation	-.380*	1	.572**
	Sig. (2-tailed)	.027		.000
	N	34	43	43

Table 51 (continued)

Compelling	Pearson Correlation	-.407*	.572**	1
	Sig. (2-tailed)	.017	.000	
	N	34	43	43

*. Correlation is significant at the 0.05 level (2-tailed).

**. Correlation is significant at the 0.01 level (2-tailed).

In the Extraversion experiment, a number of metrics were used to measure task performance: quiz score and completion time for the *quiz* task, and completion time and completion time improvement for the *math* task (see subsection 6.2.5.4, *Hypothesis 3: Task Performance*, for details). Perhaps not surprisingly, given that no differences in quiz performance were found between the conditions, there were no significant correlations between the *quiz* performance metrics and any of the subjective measures. On the contrary, some significant correlations were observed between *math* performance metrics and a number of scales/subscales.

As you may recall, two different sets of semantic differential scales were used to evaluate the tasks subjectively: for the quiz task, participants rated the robot on task appropriateness, welcome, and appeal, and for the math task, the scales of *Unobtrusiveness*, *Ease*, and *Appropriateness* were used, along with the *TLX* measure (subsection 6.2, *Evaluating Expressions of Extraversion and Introversion in a Robot as a Guide Scenario*). Additionally, the robot was rated on the Naturalness scale after both tasks were completed. Interestingly, no correlation was found between the math task performance metrics and the TLX and Ease ratings, suggesting that the perceived difficulty of the math problem was not a good indicator of performance. Completion Time Improvement for the math task was not correlated with any other math related scales. Math Completion Time, however, did correlate (negatively) at the 0.05 level with the "Easy To Tune Out" subscale of the *Unobtrusiveness* scale, and "Matched to Task" subscale of the *Appropriateness* (math) scale those participants who thought that the robot was matched to task and easy to tune out, also completed the math problem faster (**Table 52**).

Table 52: Correlations between math Completion Time metric, and *Easy To Tune Out* and *Matched to Task* subscales

		Math Completion (experimental)	Matched to Task (math)	Easy to Tune Out
Math Completion (experimental)	Pearson Correlation	1	-.393*	-.375*
	Sig. (2-tailed)		.039	.050
	N	28	28	28
Matched to Task (math)	Pearson Correlation	-.393*	1	.694**
	Sig. (2-tailed)	.039		.000
	N	28	28	28
Easy to Tune Out	Pearson Correlation	-.375*	.694**	1
	Sig. (2-tailed)	.050	.000	
	N	28	28	28

*. Correlation is significant at the 0.05 level (2-tailed).

**. Correlation is significant at the 0.01 level (2-tailed).

Completion Time Improvement for the *math* task, on the other hand, was correlated negatively with the ratings on the *Welcome* scale and most of its subscales, and on the *Appeal* scale and a number of its subscales (**Table 53** and **Table 54**, respectively); it should be noted that both scales were given to the participants after the quiz task. What this points to is that the same style of robot behavior that made the subjects feel welcome and the presentation appear more appealing was detrimental to the improvement they achieved on the experimental math task, as compared to the baseline. Interestingly, the same correlations were not observed for the Completion Time metric, with the exception of "Encouraged" subscale of the Welcome scale, for which a positive Pearson's correlation (two-tailed) of 0.383 was found, with $p < 0.044$.

Table 53: Correlations between *Completion Time Improvement* metric (*math*) and *Welcome* scale and subscales: the more welcoming the robot appeared during the quiz task, the less their math completion time improved.

		Completion Time Improvement (Math)	Desirable	Liked	Encouraged	Wanted	Welcome, Overall
Completion Time Improvement (Math)	Pearson Correlation	1	-.506**	-.536**	-.405*	-.434*	-.495**
	Sig. (2-tailed)		.006	.003	.033	.021	.007
	N	28	28	28	28	28	28
Desirable	Pearson Correlation	-.506**	1	.867**	.529**	.726**	.873**
	Sig. (2-tailed)	.006		.000	.003	.000	.000
	N	28	30	30	30	30	30
Liked	Pearson Correlation	-.536**	.867**	1	.683**	.759**	.913**
	Sig. (2-tailed)	.003	.000		.000	.000	.000
	N	28	30	30	30	30	30
Encouraged	Pearson Correlation	-.405*	.529**	.683**	1	.645**	.810**
	Sig. (2-tailed)	.033	.003	.000		.000	.000
	N	28	30	30	30	30	30
Wanted	Pearson Correlation	-.434*	.726**	.759**	.645**	1	.886**
	Sig. (2-tailed)	.021	.000	.000	.000		.000
	N	28	30	30	30	30	30
Welcome, Overall	Pearson Correlation	-.495**	.873**	.913**	.810**	.886**	1
	Sig. (2-tailed)	.007	.000	.000	.000	.000	
	N	28	30	30	30	30	30

**. Correlation is significant at the 0.01 level (2-tailed).

*. Correlation is significant at the 0.05 level (2-tailed).

Table 54: Correlations between *Completion Time Improvement* metric (*math*) and *Appeal* scale and subscales: the more appealing the participants found the robot's presentation, the less their math completion time improved.

		Completion Time Improvement (Math)	Useful	Exciting	Appeal
Completion Time Improvement (Math)	Pearson Correlation	1	-.412*	-.450*	-.476*
	Sig. (2-tailed)		.029	.016	.010
	N	28	28	28	28
Useful	Pearson Correlation	-.412*	1	.501**	.731**
	Sig. (2-tailed)	.029		.005	.000
	N	28	30	30	30
Exciting	Pearson Correlation	-.450*	.501**	1	.823**
	Sig. (2-tailed)	.016	.005		.000
	N	28	30	30	30
Appeal	Pearson Correlation	-.476*	.731**	.823**	1
	Sig. (2-tailed)	.010	.000	.000	
	N	28	30	30	30

*. Correlation is significant at the 0.05 level (2-tailed).

**. Correlation is significant at the 0.01 level (2-tailed).

Both math task performance metrics were also correlated with "Appropriate" subscale of the *Quiz Appropriateness* scale, positively with math *Completion Time*, and negatively with *Completion Time Improvement* (**Table 55**), indicating that the more appropriate the subjects found the robot for the quiz task, the worse they did on the math problem – they both took longer, and improved the completion time less. Finally, Completion Time Improvement variable had a high negative correlation with the "Lifelike" subscale of the Naturalness scale (**Table 56**) – the more lifelike the robot appeared, the less the subjects improved their math completion time.

Table 55: Correlations between math performance metrics and "Appropriate" subscale of the *Quiz Appropriateness* scale: the more appropriate the participants found the robot for the quiz task, the worse they performed on the math task.

		Math Completion (experimental)	Completion Time Improvement (Math)	Appropriate (quiz)
Math Completion (experimental)	Pearson Correlation	1	-.585**	.398*
	Sig. (2-tailed)		.001	.036
	N	28	28	28
Completion Time Improvement (Math)	Pearson Correlation	-.585**	1	-.378*
	Sig. (2-tailed)	.001		.047
	N	28	28	28
Appropriate (quiz)	Pearson Correlation	.398*	-.378*	1
	Sig. (2-tailed)	.036	.047	
	N	28	28	30

**. Correlation is significant at the 0.01 level (2-tailed).

*. Correlation is significant at the 0.05 level (2-tailed).

Table 56: Correlation between *Completion Time Improvement* (math) and "Lifelike" subscale of the *Naturalness* scale: the more lifelike the participants found the robot, the less their math completion time improved.

		Completion Time Improvement (Math)	Lifelike
Completion Time Improvement (Math)	Pearson Correlation	1	-.455*
	Sig. (2-tailed)		.015
	N	28	28
Lifelike	Pearson Correlation	-.455*	1
	Sig. (2-tailed)	.015	
	N	28	30

*. Correlation is significant at the 0.05 level (2-tailed).

7.2.3.1 Summary

Although the task of selecting or designing appropriate compliance and task performance metrics is rather challenging, every effort should be made to include such measures into evaluation of affective robotic behavior effectiveness. There are a number of reasons for this recommendation:

1. Task performance metrics provide an objective means for assessing social response to affective robots.

2. Task performance results may not necessarily be reflected in the participants' subjective perceptions and attitudes, thus they provide insights which would not otherwise be available. For example, someone who performs very well on a test may still find this test difficult and demanding, or someone who finds an obnoxious robot distracting may not necessarily perform worse. In particular, the correlations between objective task performance measures in *TAME* HRI experiments were not as obvious and straightforward as might have been expected:

    - The correlations between subjective and objective measures in the Search-and-Rescue experiment were related to how persuasive the participants found the robot's request to evacuate, but not to how understandable its behavior seemed, or how natural the robot appeared.

    - In the extraversion experiment, there was hardly any correspondence between the subjective ratings of the math task (the robot's appropriateness and unobtrusiveness, and the perceived difficulty/workload of the math problem); this finding was especially surprising with respect to the *TLX* score.

- Another interesting observation was that *Completion Time Improvement* was linked more often with the subjective measures than was the raw *Completion Time*; perhaps this indicates that the improvement variable was less susceptible to individual differences in completion time, as it was measured as a percentage.
- Finally, the link between more positive assessments of the robot during the quiz task and the decreased performance during the math task strengthens the original hypothesis that some personality traits are better suited for some tasks, but not for others.

Overall, it is important to employ both subjective and objective means of evaluation, as only together they can provide a comprehensive picture.

7.3 SUMMARY AND RECOMMENDATIONS

This chapter addressed the third subsidiary research question: "What are the metrics for evaluating affective robotic behavior?" In it, different types of evaluation measures (namely, *self-assessments*, *interviews*, *observational (behavioral) measures*, *psychophysiology measurements*, and *task performance metrics),* currently in use in the HRI research community, were presented, and their advantages and disadvantages were highlighted. The overview was followed by a discussion of three different means for evaluating affective robotic behaviors, all tested in a number of affective HRI studies: 1) established psychometric tools to measure mood and personality; 2) novel semantic differential scales, suitable for reuse in other experiments; 3) task performance metrics measuring differences in participants' task performance as a response to affective robotic expressions. Based on our experience with these measures with regards to evaluating affective robots, the following observations and recommendations can be made:

1. **Evaluating Social Response.** While evaluating affective robotic behavior in human-robot interaction experiments, it is important to remember that in many cases the effectiveness of such behavior cannot be measured directly, by evaluating a robot's performance, but rather through the social response the robot invokes in participants. For example, the benefit of robotic expressions of nervousness and anxiety is not in how well a robot can evacuate from a dangerous zone itself, but rather in how well it can persuade the humans to do so.

2. **Metrics Selection.** As measures commonly used for evaluation in the HRI domain have their advantages and disadvantages (e.g., self-assessment techniques are easy to use, but suffer from lack of objectivity), it is important to perform a cost-and-benefit analysis in order to select the metrics most appropriate for an experiment. For example, if both time and resources are at a premium, then the time- and resource-consuming nature behavioral analysis would put it at a great disadvantage; however, if the benefits of such analysis outweigh its disadvantages (if experienced coders are readily available and there is existing research connecting particular behaviors to the subject-matter of the hypotheses), then behavioral analysis would be still suitable for use. For further information on evaluation criteria for metrics selection the reader is referred to Donmez et al. [203].

3. **Use of Established Psychometric Tools.** Two established psychometric tools for measuring affect (PANAS for Positive and Negative Affect [162], and Mini-Markers for personality [163]) can be recommended as appropriate means for evaluating both subjects' and a robot's affective state and personality, based on

their successful use in a number of *TAME* HRI experiments. With respect to these tests, the following additional findings were observed:

- Subscales of the PANAS tool measure different aspects of Positive and Negative Affect. Therefore analyzing individual subscales in addition to the overall ratings would provide finer-grained results, especially in cases where a particular aspect of mood is of primary interest (e.g., nervousness, hostility, excitement, etc.).
- It was found that people are likely to project their own affective state onto the robot they are interacting with, especially if the robot does not display any discernable affect otherwise. Interestingly, this phenomenon was not observed with personality traits. This finding should be taken into account when analyzing the results of these tests, in order to differentiate the recognition of affective behaviors from people's projection of their own affective state.
- Internal consistency reliability on Mini-Markers used to asses a robot's personality was comparable, if somewhat lower than that reported for its use with human subjects [163], confirming the suitability of using this tool in HRI studies.

4. **Reusable Semantic Differential Scales.** 8 novel semantic differential scales measuring a variety of concepts were designed for this dissertation. These scales were found to have at least acceptable (over 0.7), but in most cases much higher (up to 0.942 for Appropriateness) internal consistency reliability, and therefore can be recommended for use in other HRI experiments to promote repeatability. These scales cover a variety of concepts relevant to the HRI domain, and are flexible enough to be used in a variety of scenarios and robot

tasks. For example, *Persuasiveness* scale can be applied to a robot's request, message, speech, actions, etc., and would be useful in any scenario in which a robot attempts to convince participants to perform a certain task (e.g., evacuate from a dangerous zone, or perform proscribe rehabilitative exercises). **Table 57** provide a summary of these reusable semantic differential scales, including internal consistency reliability score, number of participants tested on, and concept measured for each scale.

Table 57: Summary of Semantic Differential Scales used in *TAME* HRI experiments

Semantic Differential Scale	Reliability (Cronbach's Alpha)	Number of Participants	Concept Measured
Understandability	0.714	43	How understandable is the robot?
Persuasiveness	0.799	43	How persuasive is the robot?
Naturalness (combined)	0.803	71	How natural does the robot appear?
Appropriateness (combined)	0.942	30	How appropriate is the robot for a particular task?
Welcome	0.914	30	How welcome does the robot make participants feel?
Appeal	0.848	30	How appealing are facts (actions) given (performed) by the robot?
Unobtrusiveness	0.927	28	How unobtrusive was the robot?
Ease	0.875	29	How easy did a task performed with a robot appear?

5. **Importance of Task Performance Metrics.** It is particularly important to include task performance metrics into evaluation of affective robotic behavior effectiveness. The primary reasons for this recommendation are as follows:

- Unlike most other measurements, task performance metrics provide an objective means for assessing social response to affective robots.
- Task performance results may not necessarily be reflected in the participants' subjective perceptions and attitudes, thus they provide insights which would otherwise be unavailable.

8 CONCLUSION

Step by step the state of the art in robotics advances: visual and auditory perception achieves greater precision, communications become more reliable, mobility improves, bringing us closer and closer to the time when robots become an indelible part of our everyday lives, moving from manufacturing plants into our homes and workplace. As robots gain more autonomy and start interacting with people untrained in robotics, it becomes increasingly important for them to be able to communicate in a way easily understandable and acceptable to nurses and patients in a hospital setting, or elderly in their homes, or visitors at museums and exhibitions.

Humans are inherently social creatures, and apply social rules not only to their interactions with one another, but also to those with non-human animals, and even inanimate objects. This propensity of people to anthropomorphize certain objects has been well established by Nass and his colleagues in an extensive set of experiments [12], which showed that people treat computers as social actors, whether they recognize it or not, and that even minimal cues evoke social responses. As affect plays a vital role in human social interactions (Oatley et al. [13] call emotions “the language of human social life”), it would be beneficial for robots to be able to “speak” this language. Giving robotic systems the ability to communicate with humans affectively would not only allow people to extend their known social models to robots, but would also help robots invoke desired responses from their human interaction partners or even passers-by (e.g., assisting a stuck robot, or evacuating from a dangerous zone).

To the end of promoting more natural, satisfying and effective human-robot interaction and enhancing robotic behavior in general, an integrative framework of time-varying affective robotic behavior was designed. This psychologically inspired framework (TAME) encompasses 4 different yet interrelated affective phenomena: personality

Traits, affective Attitudes, Moods and Emotions. Traits determine consistent patterns of behavior across situations and environments and are generally time-invariant; attitudes are long-lasting and reflect likes or dislikes towards particular objects, persons, or situations; moods are subtle and relatively short in duration, biasing behavior according to favorable or unfavorable conditions; and emotions provide a fast yet short-lived response to environmental contingencies. The software architecture incorporating the TAME framework was designed as a stand-alone process to promote platform-independence and applicability to other domains and its implementation on a humanoid robot was provided. Finally, the effectiveness of affective robotic behavior was explored and evaluated in a number of human-robot interaction studies with over 100 participants.

This chapter revisits each research question posited in the beginning, and underlines the contributions, stemming primarily from the exploration of these research questions.

8.1 RESEARCH QUESTIONS

The development of the *TAME* framework followed the entire design cycle for an AI-based system, from 1) transforming psychological and cognitive science theories into a mathematical and computational representation to 2) software design and implementation on a humanoid robotic platform to 3) testing and evaluation with over 100 participants. This was done in order to answer the primary research question: "*Does integration of coordinated time-varying affective processes (namely, emotions, moods, affective attitudes and personality traits) into behavior-based robotic systems generate more effective robotic behavior from the human-robot interaction standpoint?*" This general question was explored through 3 subsidiary questions, which are summarized below.

8.1.1 SUBSIDIARY QUESTION 1

Before any questions regarding effectiveness of a system can be answered, the system has to be designed, and this is where the first research subquestion came in: "*How can the aforementioned phenomena be modeled computationally in a robotic system, both individually and relative to each other*?" In particular, the following issues were addressed through this dissertation, and explored in detail in Chapter 3, *Architectural framework*:

- *What are the psychological foundations for each of the components and their interactions?* To answer this question, a large number of theories and experimental results from the fields of psychology, cognitive science and affective computing were carefully examined, and those most appropriate for robotics selected and adapted (sections 3.2.1, 3.3.1, 3.4.1 and 3.5.1, for each of the affective phenomena).

- *How can these phenomena be represented, generated, and applied to robotic behavior?* Based on the psychological findings, the generative mechanisms were designed for each phenomenon, and multiple methods were provided for their application to robotic behavior, including continuous modification of behavioral parameters and discreet affective expressions as applied to a humanoid robot (sections 3.2.2, 3.3.2, 3.4.2, and 3.5.2, for each of the affective phenomena).

- *What are their functions, and what is their relevance for robotic behavior?* The relevant functions of affect were described as they apply to human-robot interaction (sections 3.2.1, 3.3.1, 3.4.1 and 3.5.1, for each of the affective phenomena), and a subset of them was evaluated in HRI experiments, e.g., the communicative function of fear and task adaption of extraversion (sections 6.1, *Evaluating Expressions of Negative Mood and Fear in a Search-and-Rescue*

Scenario and 6.2, *Evaluating Expressions of Extraversion and Introversion in a Robot as a Guide Scenario*).

- *What are the interactions between them that can provide additional benefit beyond that of each individual component?* In humans, the affective phenomena do not exist in isolation; on the contrary, it is the interplay between them, especially in terms of the space they occupy on the time continuum, which makes them so pervasive in human behavior. The influence of personality traits and current mood on emotion generation, of emotions on mood generation, and the manifestation of attitudes through corresponding emotions were modeled specifically in the framework. This interconnectedness creates more diverse, timely and relevant responses than each phenomenon would separately. These interactions were explored throughout Chapter 3, both from a psychological point of view and mathematically.

Chapter 3, *Architectural framework*, emphasized that the differences between the affective components in *TAME* are numerous and multi-faceted and each component provides a unique advantage. One dimension along which they differ is time, including both duration and rate of change. Emotions are the most short-lived of the four, and are fast to rise and fast to decay; moods are longer in duration and change slowly and cyclically; attitudes, once formed, last for a while and are hard to influence; and finally, traits are more or less time-invariant. Another dimension of difference is object-specificity: emotions and attitudes arise in response to a specific object or situation, whereas traits and moods are diffuse, global, and apply at all times. In general, *Traits, Emotions, Moods and Attitudes* differ in their psychological, cognitive, and behavioral functions; their antecedents and generation mechanisms; their influence on behavior and application for human-robot interaction; and their duration and changes they undergo with time. Individually, each of them has a notable yet limited potential for robotics. Together, they provide a stepping stone for promoting more natural, satisfying and effective interaction between humans and robots, as evidenced by the results of two TAME HRI studies conducted as part of this dissertation (sections 6.1, *Evaluating Expressions of Negative Mood and Fear in a Search-and-Rescue Scenario* and 6.2, *Evaluating Expressions of Extraversion and Introversion in a Robot as a Guide Scenario*).

These psychological and mathematical foundations for affective robotic behavior were then grounded by translating the theory into a particular software architecture and implementation on a physical robot (Chapter 5, *software architecture and implementation*). This step was particularly important in a field as practical as robotics, where embodiment adds another layer of complexity. The software architecture incorporating the *TAME* framework was designed as a stand-alone process to achieve platform-independence and generalizability. With an interface to connect to the system's *TAME Communication Manager* (to supply sensory data), and appropriate configuration files, this software can potentially be integrated into any robotic platform or an autonomous agent system without a substantial redesign.

The software architecture consists of: *TAME Manager* (the central module of the system), *TAME Communication Manager* (receives sensor data and passes the updated affective values to the robot), a module for each of the affective components, and *Stimuli Interpreter* (processes incoming sensory input). The *TAME Module* was incorporated into *MissionLab*, a Multiagent Mission Specification and Execution robotic software toolset, and tested on Aldebaran Robotics' Nao humanoid platform (section 5.2.2., *Integration with MissionLab and Nao Robot*).

8.1.2 SUBSIDIARY QUESTION 2

The TAME framework design and implementation provided the groundwork for addressing the second research subquestion: “What are the implications for Human-Robot Interaction? Does complex affective robotic behavior lead to more natural, effective, and satisfying interaction between humans and robots?” To evaluate the effectiveness of robot affect in HRI, a set of nonverbal affective behaviors was designed and implemented on a small humanoid, Aldebaran Robotic’s Nao robot (Chapter 5, *software architecture and implementation)*.

People are very capable of reading nonverbal affective displays; they can recognize the traits of extraversion and conscientiousness, and negative affective state from exposure as short as 5 seconds, and positive affect, neuroticism, openness and agreeableness in 20 seconds of exposure to video clips displaying interpersonal interaction [183]. Unfortunately, most robots at present, even humanoids, especially those without changeable facial features, lack the wealth of human expressive capabilities – for example, shrugging shoulders, wringing hands, or fidgeting with a pencil or clothes. Whether or not the affective cues projected by such imperfectly expressive robotic platforms would be sufficient for people to recognize them as particular expressions of affect was tested in an online survey (subsection 5.3, *Online Survey on Recognition of Affective Robotic Behavior*). The participants were asked to judge the display of negative and positive moods, emotions of joy and fear, and extraversion and introversion on a humanoid robot from a set of short video clips.

As a result, it was confirmed that people can indeed correctly recognize a humanoid robot’s affective expressions:

- The recognition rates for emotions of Joy (85%) and Fear (81%) were high and comparable to those obtained in judgments of joy and fear portrayals by human actors in movie clips;

- The Extraverted robot was rated significantly higher on the extraversion dimension that the Introverted one;
- The Positive Affect score for the robot displaying positive mood was significantly higher than that for the negative robot mood, and vice versa, its Negative Affect score was significantly lower than that of negative robot mood.

These manifestations of affect through non-verbal behavior were later used to directly address the issue of the implications robotic affect may have for human-robot interaction. To this end, two formal HRI experiments were conducted to evaluate a subset of the phenomena modeled in the *TAME* framework: Negative Mood, the emotion of Fear, and the trait of Extraversion (Chapter 6, *Evaluating Robot Affect in HRI Experiments*).

8.1.2.1 Evaluating Mood and Emotions

The goal of the first of these experiments (subsection 6.1, *Evaluating Expressions of Negative Mood and Fear in a Search-and-Rescue Scenario*) was to identify the effect the display of negative mood and fear by a humanoid robot (Aldebaran Robotic's Nao) has on participants' perception of the robot and on their compliance with the robot's request to evacuate, in the context of a mock-up search-and-rescue scenario, where the situation became progressively "dangerous". After a brief tour of a simulated site of a recent explosion, the robot asked the participants to evacuate, using both an indirect request first, and then a direct one. It was hypothesized that the participants would find the affective behavior more understandable, more persuasive and more natural, and would respond to it by 1) experiencing an increase in negative mood (becoming more nervous along with the robot) and 2) complying with the robot's request to proceed to the exit to a greater extent.

Even though the participants did not consciously recognize either Negative Affect or Fear in Nao, they did, nonetheless, react to those expressions as was hypothesized. In particular:

1. The participants found the robot's request to evacuate more **compelling**, **sincere** and **convincing** when the robot was expressing either negative mood or both negative mood and fear.
2. They complied with the robot's request to "evacuate" to a greater extent in the affective conditions:
    - The subjects were faster in complying with the robot's request to leave the "dangerous" zone (when both Negative Mood and Fear were exhibited)
    - They were more prone to respond to an indirect request to evacuate when Nao was displaying affect;
    - More of those in the affective conditions walked further towards the exit than in the case of no robot affect.
3. The participants reported feeling more nervous after interacting with the robot expressing both negative mood and fear than the one showing no affect, potentially making them more alert to any unfavorable changing in the surroundings.
4. Finally, the robot expressing both mood and emotion was rated as more "conscious", and those who found Nao appear more natural, also found its request to leave more persuasive.

8.1.2.2 Evaluating Personality Traits

Personality traits help people find their niches and excel at their jobs, and particular trait configurations may be especially suited for some tasks, but would be detrimental for others; similar advantages of personality apply to robots as well. The overall goal of the second *TAME* HRI experiment (subsection 6.2, *Evaluating Expressions of Extraversion and Introversion in a Robot as a Guide Scenario*) was to determine whether certain personality traits are better suitable for certain types of tasks, and would thus be beneficial for human-robot interaction. More specifically, the objective was to identify the effect of *Extraverted* and *Introverted personality* display by a humanoid robot, Nao, on participants' task performance in order to establish whether some traits are task-appropriate, as well as the effect on their perception of robot's appropriateness, friendliness, intrusiveness and naturalness in the context of a mock-up building demolition scenario. Participants were asked to perform two types of tasks: a quiz following a presentation given by an Extraverted or an Introverted robot, and solving a simple math problem. In addition to the task being performed with the robot serving as a guide ("experimental" tasks), the subject also performed similar "baseline" tasks before the robot was introduced. It was hypothesized that:

- Participants would find the extraverted robot more appropriate and welcoming for a task requiring gregariousness and engagement from the robot (namely, giving a presentation on the building demolition process), and the facts it presented more appealing.

- Participants would find the introverted robot more appropriate for and less distracting in a task requiring concentration from the participants (namely, solving a math problem), and the problem itself would be perceived as less demanding.

- Participants would perform better on quiz task in the Extraverted condition, and better on the math task in the Introverted condition.

The results of the experiment showed that personality traits do play a role in how suitable a robot is found for certain types of tasks. In particular:

1. The extraverted robot was found to be more suited for the presentation task:
    - The participants reported that the *Extraverted* robot in the *quiz* task made them feel as if their presence during the tour was more **liked**, **encouraged**, **wanted** and **desirable**.
    - The **facts** presented by the *Extraverted* robot appeared more **fun**, **exciting** and **interesting** than those given by the *Introverted* robot.
2. *Introverted* Robot was found to be more suited for the *math* task:
    - The *Introverted* robot was found to be more **appropriate**, **right for task**, **well-suited**, **proper** and **matched to task** than the *Extraverted* one with regards to the **math** task.
    - The *Introverted* robot was also rated as **easier to tune out**, **quieter**, **less demanding and offensive**, and better at **minding its own business** than the *Extraverted* one.
    - The math task was perceived as less **demanding** overall (based on TLX scores) when it was supervised by the *Introverted* robot, and in particular it was rated **lower** on *Temporal Demand*, (concern over) *Performance* and *Frustration*; also, those who perceived the robot as **more appropriate** for the *math* task and **unobtrusive** also found the problem itself **less demanding.**

- When the *Introverted* supervised the *math* task, the **task time completion** between the baseline and experimental task **improved** by a **higher** percentage.

8.1.3 SUBSIDIARY QUESTION 3

Unlike other fields of robotics, e.g., vision or gait control, where the goals, tasks and measures are straightforward and objective, the advantages of affect are much harder to quantify. In the case of affective robots, it is not the robot performance per se that needs to be evaluated, but rather the social response the robots invoke in people they interact with. The last subsidiary question addressed the issue of measuring this social response: "*What are the metrics for evaluating affective robotic behavior?"*, and was discussed in Chapter 7, *Measuring social response to affective robots*.

Based on our experience with a variety of measures with regards to evaluating affective robots (Chapter 4, *Exploratory Experimental Study*, Chapter 6, *Evaluating Robot Affect in HRI Experiments*, and section 5.3, *Online Survey on Recognition of Affective Robotic Behavior*), a number of recommendations can be made:

1. Two established psychometric tools for measuring affect (PANAS for Positive and Negative Affect [162], and Mini-Markers for personality [163]) can be recommended as appropriate means for evaluating both subjects' and a robot's affective state and personality (subsection 7.2.1, *Established Psychometric Measurement Tools*), based on their successful use in HRI studies with over 100 participants. Internal consistency reliability on Mini-Markers used to asses a robot's personality was comparable, if somewhat lower than that reported for its use with human subjects [163]. A number of further recommendations can be made regarding the use of these tests:

- Subscales of the PANAS tool measure different aspects of Positive and Negative Affect. Therefore analyzing individual subscales in addition to the overall ratings would provide finer-grained results;

- It was found that people are likely to project their own affective state, but not their personality onto the robot they are interacting with. This finding should be taken into account when analyzing the results of these tests, in order to differentiate the recognition of affective behaviors from people's projection of their own affective state.

2. 8 novel semantic differential scales measuring a variety of concepts were designed for this dissertation and employed in *TAME* experiments (Chapter 6, *Evaluating Robot Affect in HRI Experiments*). Given their acceptable internal consistency reliability (subsection 7.2.2, *Semantic Differential Scales*), their applicability to a wide variety of HRI scenarios, and their reusability, these scales can be recommended for use in HRI experiments. In particular, these self-assessment metrics which provide participants' ratings on a number of concepts relevant to the HRI domain: Understandability, Persuasiveness, Naturalness, Welcome, Appeal, Unobtrusiveness and Ease.

3. Every effort should be made to include task performance metrics into evaluation of affective robotic behavior effectiveness, even though selection and/or design of such metrics with respect to measuring social response is particularly challenging. The primary reasons for this recommendation are:

   - Unlike most other measurements, task performance metrics provide an objective means for assessing social response to affective robots.

- Task performance results may not necessarily be reflected in the participants' subjective perceptions and attitudes, thus they provide insights which would otherwise be unavailable (section 7.2.3, *Task Performance Metrics*).

Examples of using task performance metrics to evaluate social response to affective robots can be found in section 6.1, *Evaluating Expressions of Negative Mood and Fear in a Search-and-Rescue Scenario, and 6.2, Evaluating Expressions of Extraversion and Introversion in a Robot as a Guide Scenario*. In particular, the metrics used were request compliance (measured as how fast, how soon, and to what extent participants complied with a robot's request) and task performance (correctness, task completion time, and improvement in completion time between baseline and experimental tasks).

4. Finally, while evaluating affective robotic behavior in human-robot interaction experiments, it is important to remember that in many cases the effectiveness of such behavior cannot be measured directly, by evaluating a robot's performance, but rather through the social response the robot invokes in participants. This recommendation should be taking into consideration when selecting and/or designing measures to be used in HRI experiments assessing robot affect.

8.2 CONTRIBUTIONS

The contributions to the field of robotics produced as a result of this dissertation stem directly from the exploration of the aforementioned research issues:

- ***An integrative framework to augment behavior-based robotic systems with a variety of time-varying affective processes, namely emotions, moods, attitudes and personality***. Not only was the software design incorporating the *TAME* framework and its implementation on a humanoid robot supported by a grant from a successful commercial company, Samsung, but it was also integrated into a robotic planner developed independently from this dissertation work [204]. The framework design, initial evaluation in an exploratory longitudinal study, software architecture and implementation are described in Chapters 3-5.

- ***Means for affective communication to afford more effective human-robot interaction***. Together, the generative framework (Chapter 3), the design and implementation of specific affective expressions on a humanoid robot (Chapter 5), and the successful evaluation (Chapters 4, 6 and 7) present a tool set for designing affective robotic behaviors with the goal of promoting more natural, satisfying and effective interaction between humans and robots.
- ***Metrics for evaluating effectiveness of affective robotic behavior.*** The measures examined or developed in this dissertation: the psychometric tests of mood and personality, semantic differential scales for measuring a number of HRI-relevant concepts, and compliance and task performance metrics provide a collection of tools which can be employed to evaluate social response to affective robots (Chapter 7, *Measuring social response to affective robots*).

To conclude, we believe that multi-faceted robotic affect can have far-reaching practical benefits for human-robot interaction, from making people feel more welcome where gregariousness is expected to making unobtrusive partners for problem solving tasks to saving people's lives in dangerous situations. If measurable effect of affective robotic behavior could be obtained as a result of short interactions in experimental settings, the impact is likely to be greater in prolonged, everyday exposure to robots.

APPENDIX A: CONSENT FORM FOR THE AIBO EXPLORATORY EXPERIMENTAL STUDY

Georgia Institute of Technology
Consent to Be a Research Participant

1. Title of Research Project
Empirical Study for Designing Human-Robot Collaborative Teams.

2. Principal Investigators
Ronald C. Arkin and Christine M. Mitchell

3. Purpose of Research
You are being asked to volunteer for a research study. This study is a GVU, Georgia Institute of Technology, sponsored exploratory empirical study conducted as a part of research for GVU seed grant (College of Computing, GA Institute of Technology). The long-term goal of the research project is to create, demonstrate, and evaluate methods to design human-robot teams that collaborate effectively, and to build a framework of affective robotic behavior to facilitate the collaboration. The framework is being developed within Mission Lab programming environment (software for Multiagent Robotics Mission Specification and Control), and will include a number of affective components, such as personality traits, emotions, moods and attitudes. The purpose of the empirical study is to identify affective attributes of effective human-robot collaboration that the framework can incorporate and explore human reaction to the interaction with a robot with respect to ease and pleasantness of interaction. In particular, we would like to explore the change in human attitude over time and the idea of companionship between humans and robots.

4. Procedure
You are asked to participate in the experimental study aimed at identifying key affective aspects of human-robot interaction. It involves interacting (playing, petting, addressing) with a Sony entertainment robotic dog (AIBO ERS-210A) for a total of 4 sessions, 20-45 minutes each. The sessions will be conducted in TSRB, room 236, or the Mobile Robot Laboratory, and your interactions with the robot will be recorded via a video/audio recorder. You will be also asked to answer various written and/or oral questions that will help assess your attitude towards the robot, and the quality of interaction. A more detailed description of the procedure will be provided by the administrator prior to the start of the session.

5. Foreseeable risks or discomforts:
This research involves minimal risk. The risks involved are no greater than those involved in daily activities such as checking your e-mail or playing with a pet.

6. Benefits
There is no direct benefit to you by participating, however, you will have an opportunity of interacting with an entertainment robotic dog, and expand your knowledge about robotic pets during the session.

7. Compensation/Costs
You will be paid for your participation in the study: $10 for each of the first three sessions, and $30 for the last session, for a total of $60.

8. Confidentiality
The following procedures will be followed to keep your personal information confidential in this study: The data that is collected about you will be kept private to the extent allowed by law. To protect your privacy, your records will be kept under a code number rather than by name. Your records will be kept in locked files and only study staff will be allowed to look at them. Your name and any other fact that might point to you will not appear when results of this study are presented or published. The video/audio tapes of you will be stored for archival purposes for the duration of two years at Mobile Robot Lab. Only the experimenters and the parties delineated below will have access to the tapes. The tapes will be analyzed with respect to your attitude towards the robotic dog and the analysis will be included in the results of the study; however, neither your name nor your appearance will be disclosed in the results. Any additional use of the tapes (e.g., playing a video at a conference) will require a written consent from you.

To make sure that this research is being carried out in the proper way, the Georgia Institute of Technology IRB will review study records. The sponsor of this study, GVU, Georgia Institute of Technology, has the right to review study records as well. Again, your privacy will be protected to the extent allowed by law

9. Injury/Adverse Reactions

Reports of injury or reaction should be made to Ronald C. Arkin at (404) 894 – 8209. Neither the Georgia Institute of Technology nor the principal investigator has made provision for payment of costs associated with any injury resulting from participation in this study.

10. Contact Person

If you have questions about the research, call or write to the principal investigators at:

Ronald C. Arkin
College of Computing
Georgia Institute of Technology
Atlanta, GA 30332-0280
(404) 894 – 8209 (voice)
(404) 894 – 0957 (fax)
arkin@cc.gatech.edu

Christine M. Mitchell
Industrial & Systems Engineering
765 Ferst Dr.
Georgia Institute of Technology
Atlanta, GA 30332-0205
404-894-4321 (voice)
404-385-0357 (fax)
cm@chmsr.gatech.edu

11. Voluntary Participation/Withdrawal

You have rights as a research volunteer. Taking part in this study is completely voluntary. If you do not take part, there is no penalty. You may stop taking part in this study at any time with no penalty. If you have any questions about your rights as a research volunteer, call or write:

The Institutional Review Board,
Office of Research Compliance
505 Tenth Street, 3rd Floor
Georgia Institute of Technology
Atlanta, Georgia 30332-0420
(404) 894 – 6942 (voice)
(404) 894 – 00864 (fax)
IRB@gatech.edu

12. Signatures

I have read the information above. The researchers have answered all my questions to my satisfaction. They gave me a copy of this form. I consent to take part in this study.

Participant's Signature: ____________________ Date: ______ Time: ______

Person Obtaining Consent: ____________________ Date: ______ Time: ______

APPENDIX B: SCRIPT FOR EXPERIMENT ADMINISTRATORS

Aibo Study 2004 Script

Session 1:

Hi, my name is ______ and I'd like to thank you for participating in this study. First off, I will be reading this script to ensure that everyone who participates gets the same directions.

In this study, we are examining how people and robots perform certain tasks. During each session, you will be given the opportunity to interact with Aibo, a Sony entertainment robot dog. Each of the four sessions will last between 25 and 45 minutes including pre and post questionnaires. We will also teach you new ways to interact with the robotic dog and give you new tasks to perform. As this robot is a toy made for kids, it will pose minimal harm to you.

You will receive a total of $60 dollars for you participation in this study - $10 for each of the first three sessions and $30 for the last session. You may stop at any point, but you will only be paid for the sessions you have already completed.

Also, while you are doing these tasks, we will be videotaping these sessions. The camera is located here *(point to location)*. The videotapes will help us analyze your interactions with the robotic dog. In case you are concerned about privacy, we will remove all connections between your name and the data so the only way someone can know it's you in the videotapes is if they recognize you. The questionnaires and the video tapes will also be stored in a securely locked cabinet at Georgia Tech. Only members of the research team will ever see these videos. However, if you don't mind us showing these videotapes to other researchers, for example, at a conference, there is a form you can fill out at the end of the study to give us permission to show it to people outside of the research team.

At this point, I'd like to ask if you have any questions.

Okay, please take a moment to read over and sign the consent form. One of the copies is for you. Also, in order to get paid, you need to fill out this payment form.

Before I introduce you to the robotic dog, there are two questionnaires I need you to fill out. The first one is a demographics questionnaire. Please fill out the questionnaire and tell me when you are done (hand out the demographics survey). Thank you. The second questionnaire is about your personality. Don't worry. This is not about you. We simply need this information to analyze your interactions with the robot. Please fill out the questionnaire and tell me when you are done *(hand out MINI MARKERS).*

We are just about to begin – I'll go ahead and start the camera and the robot, and we'll be ready in a second. *(start camera and robot)* Alright! Now we are ready to begin. If at any point you'd like to stop interacting with the robot, you can tell it to either stop, or go play. To make the robot stop, you can give the verbal command: "Stop". To make it wander around without needing your interaction, you can say the command: "Go Play", and the robot will start wandering around on its own. It may sometimes fall down and will need your help getting back up. In such a case, please pick it up like this *(show participant)*. Also try to keep the robotic dog inside the green field so we can capture everything on camera, and if it's about to run into a hard surface, please turn it around. Feel free to pick up the robotic dog if you need to help it see. I am now going to teach you the first command. The robot will only recognize the commands I teach you, in addition to "Stop" and "Go Play". When you talk, especially when you give the commands, please speak loudly – this is very important.

For the first task, I would like you to tell the robot to follow the ball. Please pick up the pink ball from the box over there. Now, say the command, "follow the ball". Every time you want the robot to start performing a command, you need to make sure that you say that command. At any time, you can tell it to stop or go play. Tell me when you think you have succeeded in making the robot follow the ball. *(After the task is completed)* Now that you are done with the task, I'd like you to make the robotic dog perform the same task at least three times in a row, but in order to separate these commands from each other, please make the robot either stop or wander around in between the "Follow the Ball" commands. Also, please make sure you hide the pink ball between the commands. When you are done, tell it to either Stop or Go Play.

Are you ready to learn the next command? Okay. In addition to following the ball, you can tell the robot to kick the ball. Say the command, "kick the ball" for the robot to do so. Tell me when you think you have succeeded in this task. *(wait until participant done with task)* As with the previous task, I'd like you to repeat the task at least 3 times, making it either stop or go play in between.

You can continue interacting with the robot if you'd like, using any of the commands you've learned – just let me know when you are done. *(wait for "done")* Okay, I have one more questionnaire for you to fill out and then this session is over *(hand out PANAS-T;* fill out the payment form for this session, schedule the next session if not scheduled yet).

Session 2:

(start camera) Welcome back! In this session, we will teach you two more commands. While you are learning these new commands, you may use any of the commands you learned in the previous session. Just to recap, the commands you learned were: Follow the ball, and Kick the Ball. You can also tell the dog to Stop or Go Play at any time, just as in the previous session. I'll go ahead and start the robot now, and we'll begin. *(start robot)*

In this session you will need to make the robotic dog follow you and come to you. In order for the robot to recognize you, you will need to hold these flowers – they are a representation of you for the robot. Please remember, that you need to hide the ball or the flowers between the commands you give the dog – you can use this box to put the props away, if you'd like to – or hide them behind you so that the robotic dog won't see them. Are you ready for the first command today? For the first command, you will ask the robot to follow you. Do so by saying the command "follow me". Remember to hold the flowers if you want the robot to recognize you. Tell me when you think you have succeeded in this task. *(After the subject is done with the task)* Now that you are done with the task, I'd like you to make the robotic dog perform the same task at least three times in a row, with making it either stop or wander around in between, or you can use any of the commands learned previously. Please remember to select the appropriate prop: the ball for the ball commands, and the flowers for the commands that refer to you.

Are you ready to learn the next command? The next command will tell the robot to come to you. Say the command, "come to me" in order to do this. You will need to use the flowers again. Tell me when you think you have succeeded in this task. *(wait until participant done with task)* As with the previous task, I'd like you to repeat the task at least 3 times, making it either stop or wander around in between, as well as using any commands from the last session.

Now, please take at least five minutes to practice the commands from this and the previous sessions, using the commands in any order. Please remember to pick up and put away the appropriate prop while doing so.

Are you done playing with the robotic dog? Okay, I have one more questionnaire for you to fill out and then this session is over *(hand out PANAS-T).* Let's schedule your next session. Thank you for coming. *(Fill out the payment form for this session, schedule the next session if not scheduled yet).*

Session 3:

(start camera) Welcome back! In this session, we will teach you one final command. While you are learning this new command, you may use any of the commands you learned in the previous sessions. Just to recap, the commands you learned were: Follow

the ball, Kick the Ball, Follow Me and Come to Me. You can also tell the dog to Stop or Go Play at any time, just as in the previous sessions. I'll go ahead and start the robot now *(start robot)*.

In this session an intruder robot will periodically come into the room. Imagine that this intruder is very annoying, and you'd like the robotic dog to drive it away, and so protect you from it. You may play with the robotic dog as you wish until you see the intruder. To help draw your attention, just before the intruder robot appears, it will make this sound. *(connect to amigobot)* When you hear this sound, please hide all other props, as they will distract the robotic dog. In general, please remember that you need to hide the ball or the flowers between the commands you give the robotic dog – you can use this box to put the props away, if you'd like to – or hide them behind you so that the robotic dog won't see them. When you do see the intruder, say the command "Sic' em" *("Attack" if sic'em is not understood)* to make the robotic dog attack the intruder. After the robot attacks the intruder, the intruder will withdraw. *(After the subject is done with the task, and amigobot is gone)*: now that you are done with the task, please continue interacting with the dog until the intruder robot appears again – this will happen 3 more times during this session. While waiting for the intruder, you can make the robot stop, wander around, or perform any of the other commands learned previously. Please remember to select the appropriate prop: the ball for the ball commands, and the flowers for the commands that refer to you. *(Bring out the amigobot every 4 minutes, not counting the time it takes to bring the amigobot out and back)*.

You can continue interacting with the robot if you'd like, using any of the commands you know – just let me know when you are done. Are you done interacting with the dog? Okay, I have one more questionnaire for you to fill out and then this session is over *(hand out PANAS-T)*. Let's schedule your next session. Thank you for coming. *(Fill out the payment form for this session, schedule the next session if not scheduled yet).*

Session 4:

Welcome back! In this session, we will not teach you any more commands – this is your opportunity to interact with the robot. You are free to interact with the robot and use any of the commands you have already learned. The flowers and the ball are both in this box. Just to recap, the commands you learned were: Follow the ball, Kick the Ball, Follow Me, Come to Me, and Sic' em. As soon as you see the intruder – or hear this sound *(connect to the amigobot)*, please hide the ball or the flowers, and sic the intruder robot. You can also tell the dog to Stop or Go Play at any time, just as in the previous sessions. Please remember to select the appropriate prop: the ball for the ball commands, and the flowers for the commands that refer to you. Make sure you put the prop away in this box before you take another one out, or hide it so that the robotic dog doesn't see it. The intruder robot will appear every now and then, so you can tell the robot to sic'em at that time. I'll go ahead and start the robot now.

Please interact with the robot for at least 15 minutes – there's a clock on that desk. Feel free to interact with the robot longer – just let me know when you are done so we can complete the experiment.

(turn off camera when the user says "done") Okay, I have a few questionnaires for you to fill out and then this experiment is over. Again, here is the mood questionnaire *(hand out PANAS-T)*. Thank you. Here is the next questionnaire. You might remember filling out a similar one in the first session. This time, however, answer the questions in regards to the robot instead of yourself. *(hand out MINI-Markers for dog)*. Finally, here is the last questionnaire *(hand out post-questionnaire).* There's one last thing I need to ask you. We would like to get the permission to use portions of your video footage at conferences, workshops, and other research gatherings. If you don't mind, then please sign this form. (*Hand out the video release form. Answer any questions if asked)*. The experiment is now over. Thank you for participating.

APPENDIX C: DEMOGRAPHICS QUESTIONNAIRE

Subject # ______ Session #_________

Demographics Questionnaire

1. What is your gender?
   a.) Female
   b.) Male

2. What is your age?
   a.) Under 20 years old
   b.) Between 20 and 30
   c.) Between 30 and 40
   d.) Between 40 and 50
   e.) 50 or older

3. What is the highest level of education you've achieved? What was your major?
   a.) High School
   b.) Bachelor's __________________
   c.) Master's __________________
   d.) Ph.D. __________________
   e.) Currently working on my Ph.D. _________________
   f.) Other _______________________

4. What is your race/national origin?

   a.) White
   b.) African American
   c.) Asian/Pacific Islander
   d.) Hispanic
   e.) Other

5. Have you/your family/your roommate ever owned any pets? How many? Please circle all that apply.

   a.) Dog ____
   b.) Cat ____
   c.) Fish ____
   d.) Birds ____
   e.) Other (Please specify)
   f.) None

6. What is the longest you've had a pet for?

   a.) Under 3 months
   b.) 3-6 months
   c.) 7-12 months
   d.) 13 months - 3 years
   e.) Longer than 3 years
   f.) Never had one

Subject # ______ Session #_________

7. Do you think any of your pets have (had) a personality? Please circle the number above your response.

1	2	3	4	5
Of course not!	No, I don't think so	Maybe	Yes	Yes, a very vivid one

8. What is your level of computer experience?

   a.) None: Never used a computer before
   b.) Limited: Occasionally use a computer for tasks like e-mail, web browsing or word processing
   c.) User Level: Regularly use a computer for tasks like e-mail, web browsing or word processing
   d.) Advanced User: Have downloaded and installed at least one program from the Internet
   e.) Programmer Level: Some programming language or network administration experience
   f.) Advanced Programmer: Extensive training or experience in multiple programming languages

9. Have you ever interacted with robots? Please circle all that apply.

   a.) Very limited interaction
   b.) Interaction experience with industrial robots
   c.) Interaction experience with mobile robots
   d.) Interaction experience with entertainment/educational robots
   e.) Interaction experience with Sony entertainment robot Aibo
   f.) Never

APPENDIX D: POST QUESTIONNAIRE FOR AIBO STUDY

Subject # _________ Session # ____________

Please reflect back on your interaction with the robotic dog as you consider the statements below. Please circle the number above the phrase that most closely fits the extent to which you agree or disagree with each statement.

1. It was easy to get the robotic dog to perform commands.

1	2	3	4	5
Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree

2. It was easy to understand whether the dog was performing the command or not.

1	2	3	4	5
Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree

3. The robotic dog showed emotional expressions.

1	2	3	4	5
Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree

If you answered "Agree" or "Strongly Agree" to question 3, please answer the following questions (3a and 3b). Otherwise, go to question 4.

3a. Emotional expressions exhibited by the dog made the interaction more enjoyable.

1	2	3	4	5
Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree

3b. Emotional expressions exhibited by the dog made the interaction easier.

1	2	3	4	5
Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree

4. The robotic dog had a personality of its own.

1	2	3	4	5
Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree

Subject # ________ Session # ____________

If you answered "Agree" or "Strongly Agree" to question 4, please answer the following question (4a). Otherwise, go to question 5.

4a. I enjoyed interacting with the robot, partly because it possessed some personality.

1	2	3	4	5
Strongly Disagree	**Disagree**	**Neutral**	**Agree**	**Strongly Agree**

5. With every session, I was getting more attached to the dog.

1	2	3	4	5
Strongly Disagree	**Disagree**	**Neutral**	**Agree**	**Strongly Agree**

6. Overall, I enjoyed the interaction with the robotic dog.

1	2	3	4	5
Strongly Disagree	**Disagree**	**Neutral**	**Agree**	**Strongly Agree**

7. Please use the space below (attach additional sheets if needed) to describe your interactions with the robotic dog. Specifically, did the dog seem to have a personality? If so, what kind of personality? Also, describe any emotional states that you think the dog exhibited during your interaction. Please describe your own state during the interaction: e.g., entertained, bored, curious, etc. Did your attitude change to the robotic dog throughout the sessions? How? Finally, would you prefer robots that interact with humans to express some emotion and personality? Why?

Thank you very much for your participation!

APPENDIX E: A BRIEF BIG FIVE PERSONALITY QUESTIONNAIRE FOR PARTICIPANT

Subject # _________ Session # __________

How Accurately Can You Describe Yourself?

Please use this list of common human traits to describe yourself as accurately as possible. Describe yourself as you see yourself at the present time, not as you wish to be in the future. Describe yourself as you are generally or typically, as compared with other persons you know of the same sex and of roughly your same age. Before each trait, please write a number indicating how accurately that trait describes you, using the following rating scale:

Extremely Inaccurate – 1

Very Inaccurate – 2

Moderately Inaccurate – 3

Slightly Inaccurate - 4

Neither Inaccurate or Accurate - 5

Slightly Accurate – 6

Moderately Accurate – 7

Very Accurate – 8

Extremely Accurate - 9

____ Bashful	____ Energetic	____ Moody	____ Systematic
____ Bold	____ Envious	____ Organized	____ Talkative
____ Careless	____ Extraverted	____ Philosophical	____ Temperamental
____ Cold	____ Fretful	____ Practical	____ Touchy
____ Complex	____ Harsh	____ Quiet	____ Uncreative
____ Cooperative	____ Imaginative	____ Relaxed	____ Unenvious
____ Creative	____ Inefficient	____ Rude	____ Unintellectual
____ Deep	____ Intellectual	____ Shy	____ Unsympathetic
____ Disorganized	____ Jealous	____ Sloppy	____ Warm
____ Efficient	____ Kind	____ Sympathetic	____ Withdrawn

Bashful: socially shy or timid; diffident; self-conscious.
Bold: fearless before danger; intrepid; impudent; presumptuous.
Careless: Marked by lack of attention, consideration, forethought or thoroughness; not careful.
Cold: Marked by a lack of the warmth of normal human emotion, friendliness, or compassion.
Complex: Hard to separate, analyze, or solve; complicated.
Cooperative: marked by a willingness and ability to work with others.
Creative: Characterized by originality and expressiveness; imaginative.
Deep: Of penetrating intellect; wise.
Disorganized: To be into utter disorder; disarrange.
Efficient: Exhibiting a high ratio of output to input; effective.
Energetic: Operating with or marked by vigor or effect; vigorous.
Envious: Painfully desirous of another's advantages; jealous; covetous.
Extraverted: a gregarious and unreserved person; outgoing.
Fretful: Marked by worry and distress; inclined to be vexed or troubled.
Harsh: Unpleasantly stern; severe.
Imaginative: Created by, indicative of, or characterized by imagination; having no truth; false.
Inefficient: Wasteful of time, energy, or materials; lacking the ability or skill to perform effectively; incompetent:
Intellectual: Having or showing intellect, especially to a high degree; intelligent
Jealous: Fearful or wary of being supplanted; apprehensive of losing affection or position; resentful or bitter in rivalry; envious.
Kind: Of a friendly, generous, or warm-hearted nature; considerate.
Moody: Subject to depression or moods; expressive of a mood; temperamental.
Organized: Methodical and efficient in arrangement or function; orderly.
Philosophical: Characteristic of a philosopher, as in equanimity, enlightenment, and wisdom.
Practical: Concerned with actual use or practice; useful.
Quiet: Restrained in style; understated; making little or no noise.
Relaxed: To be less restrained or tense; easy and informal in manner.
Rude: Ill-mannered; discourteous; uncouth.
Shy: Marked by reserve or diffidence; reserved; wary.
Sloppy: Marked by a lack of neatness or order; untidy.
Sympathetic: Expressing or feeling compassion or friendly fellow feelings.
Systematic: Characterized by order and planning; orderly.
Talkative: Full of trivial conversation; loquacious; garrulous; voluble.
Temperamental: Marked by excessive sensitivity and impulsive changes of mood.
Touchy: marked by readiness to take offense on slight provocation; sensitive.
Uncreative: opposite of creative (see above)
Unenvious: opposite of envious (see above)
Unintellectual: opposite of intellectual (see above)
Unsympathetic: opposite of sympathetic (see above)
Warm: Marked by or revealing friendliness or sincerity; loving; kind.
Withdrawn: Not friendly or sociable; aloof; detached; emotionally unresponsive.

APPENDIX F: A BRIEF BIG FIVE PERSONALITY QUESTIONNAIRE FOR ROBOTIC DOG

Subject # __________ Session # ____________

How Accurately Can You Describe The Robotic Dog?

Please use this list of common human traits to describe the robotic dog as accurately as possible. Describe the robotic dog as you see it, based on your previous interactions. Before each trait, please write a number indicating how accurately that trait describes the robotic dog, using the following rating scale:

Extremely Inaccurate ……………… 1

Very Inaccurate …………………….. 2

Moderately Inaccurate …………….. 3

Slightly Inaccurate ………………….4

Neither Inaccurate or Accurate ……. 5

Slightly Accurate ………………….. 6

Moderately Accurate ………………. 7

Very Accurate ……………………... 8

Extremely Accurate ………………...9

_____ Bashful	_____ Energetic	_____ Moody	_____ Systematic
_____ Bold	_____ Envious	_____ Organized	_____ Talkative
_____ Careless	_____ Extraverted	_____ Philosophical	_____ Temperamental
_____ Cold	_____ Fretful	_____ Practical	_____ Touchy
_____ Complex	_____ Harsh	_____ Quiet	_____ Uncreative
_____ Cooperative	_____ Imaginative	_____ Relaxed	_____ Unenvious
_____ Creative	_____ Inefficient	_____ Rude	_____ Unintellectual
_____ Deep	_____ Intellectual	_____ Shy	_____ Unsympathetic
_____ Disorganized	_____ Jealous	_____ Sloppy	_____ Warm
_____ Efficient	_____ Kind	_____ Sympathetic	_____ Withdrawn

Bashful: socially shy or timid; diffident; self-conscious.
Bold: fearless before danger; intrepid; impudent; presumptuous.
Careless: Marked by lack of attention, consideration, forethought or thoroughness; not careful.
Cold: Marked by a lack of the warmth of normal human emotion, friendliness, or compassion.
Complex: Hard to separate, analyze, or solve; complicated.
Cooperative: marked by a willingness and ability to work with others.
Creative: Characterized by originality and expressiveness; imaginative.
Deep: Of penetrating intellect; wise.
Disorganized: To be into utter disorder; disarrange.
Efficient: Exhibiting a high ratio of output to input; effective.
Energetic: Operating with or marked by vigor or effect; vigorous.
Envious: Painfully desirous of another's advantages; jealous; covetous.
Extraverted: a gregarious and unreserved person; outgoing.
Fretful: Marked by worry and distress; inclined to be vexed or troubled.
Harsh: Unpleasantly stern; severe.
Imaginative: Created by, indicative of, or characterized by imagination; having no truth; false.
Inefficient: Wasteful of time, energy, or materials; lacking the ability or skill to perform effectively; incompetent:
Intellectual: Having or showing intellect, especially to a high degree; intelligent
Jealous: Fearful or wary of being supplanted; apprehensive of losing affection or position; resentful or bitter in rivalry; envious.
Kind: Of a friendly, generous, or warm-hearted nature; considerate.
Moody: Subject to depression or moods; expressive of a mood; temperamental.
Organized: Methodical and efficient in arrangement or function; orderly.
Philosophical: Characteristic of a philosopher, as in equanimity, enlightenment, and wisdom.
Practical: Concerned with actual use or practice; useful.
Quiet: Restrained in style; understated; making little or no noise.
Relaxed: To be less restrained or tense; easy and informal in manner.
Rude: Ill-mannered; discourteous; uncouth.
Shy: Marked by reserve or diffidence; reserved; wary.
Sloppy: Marked by a lack of neatness or order; untidy.
Sympathetic: Expressing or feeling compassion or friendly fellow feelings.
Systematic: Characterized by order and planning; orderly.
Talkative: Full of trivial conversation; loquacious; garrulous; voluble.
Temperamental: Marked by excessive sensitivity and impulsive changes of mood.
Touchy: marked by readiness to take offense on slight provocation; sensitive.
Uncreative: opposite of creative (see above)
Unenvious: opposite of envious (see above)
Unintellectual: opposite of intellectual (see above)
Unsympathetic: opposite of sympathetic (see above)
Warm: Marked by or revealing friendliness or sincerity; loving; kind.
Withdrawn: Not friendly or sociable; aloof; detached; emotionally unresponsive.

APPENDIX G: BRIEF MEASURE OF POSITIVE AND NEGATIVE AFFECT (MOOD)

PANAS-T
Circle the answer that best describes the extent to which you are experiencing each of the feelings or emotions below *right now*.

1. interested	*very slightly or not at all*	*a little*	*moderately*	*quite a bit*	*extremely*
2. distressed	*very slightly or not at all*	*a little*	*moderately*	*quite a bit*	*extremely*
3. excited	*very slightly or not at all*	*a little*	*moderately*	*quite a bit*	*extremely*
4. upset	*very slightly or not at all*	*a little*	*moderately*	*quite a bit*	*extremely*
5. strong	*very slightly or not at all*	*a little*	*moderately*	*quite a bit*	*extremely*
6. guilty	*very slightly or not at all*	*a little*	*moderately*	*quite a bit*	*extremely*
7. scared	*very slightly or not at all*	*a little*	*moderately*	*quite a bit*	*extremely*
8. hostile	*very slightly or not at all*	*a little*	*moderately*	*quite a bit*	*extremely*
9. depressed	*very slightly or not at all*	*a little*	*moderately*	*quite a bit*	*extremely*
10. enthusiastic	*very slightly or not at all*	*a little*	*moderately*	*quite a bit*	*extremely*
11. proud	*very slightly or not at all*	*a little*	*moderately*	*quite a bit*	*extremely*
12. irritable	*very slightly or not at all*	*a little*	*moderately*	*quite a bit*	*extremely*

Subject #: ___________ Session #: _____________

13. alert	*very slightly or not at all*	*a little*	*moderately*	*quite a bit*	*extremely*
14. ashamed	*very slightly or not at all*	*a little*	*moderately*	*quite a bit*	*extremely*
15. inspired	*very slightly or not at all*	*a little*	*moderately*	*quite a bit*	*extremely*
16. happy	*very slightly or not at all*	*a little*	*moderately*	*quite a bit*	*extremely*
17. determined	*very slightly or not at all*	*a little*	*moderately*	*quite a bit*	*extremely*
18. attentive	*very slightly or not at all*	*a little*	*moderately*	*quite a bit*	*extremely*
19. jittery	*very slightly or not at all*	*a little*	*moderately*	*quite a bit*	*extremely*
20. nervous	*very slightly or not at all*	*a little*	*moderately*	*quite a bit*	*extremely*
21. active	*very slightly or not at all*	*a little*	*moderately*	*quite a bit*	*extremely*
22. afraid	*very slightly or not at all*	*a little*	*moderately*	*quite a bit*	*extremely*

APPENDIX H: FREE-RESPONSE QUESTIONNAIRE SUMMARY

The questions in the free-response portion of the questionnaire were as follows:

> Please use the space below (attach additional sheets if needed) to describe your interactions with the robotic dog. Specifically, did the dog seem to have a personality? If so, what kind of personality? Also, describe any emotional states that you think the dog exhibited during your interaction. Please describe your own state during the interaction: e.g., entertained, bored, curious, etc. Did your attitude to the robotic dog change throughout the sessions? How? Finally, would you prefer robots that interact with humans to express some emotion and personality? Why?

The obtained data were summarized by sub-question and by condition in **Table 58**. The subdivision of the comments into Aibo's personality and Aibo's emotions is to an extent arbitrary, as it was clear from the responses that the participants did not have a clear understanding of the differences between personality and emotions, and seemed to use those terms interchangeably. Also, not everyone answered all of the sub-questions, therefore the total number of responses for each of the sub-questions does not add up to the total number of subjects. Each bullet point represents a response from one person; where possible, the responses are verbatim, otherwise, they are summarized as close to the original as possible (e.g., 1st person singular was changed into 3rd person singular in the table).

Table 58: Summary of responses to the free-form portion of the post questionnaire

Sub-Question	Condition A: Non-Emotional	Condition B: Emotional
Aibo's Personality	- Mechanical - body language showed courage when facing the intruder; lacked subtle facial expressions and sound variations that people read personality by in pets - warm personality, like that of a puppy - friendly, playful, slow, persistent - finicky - quasi-personality: loyal, commanding, bold, but devoid of true affection; when the dog would stop and wait for the next command, the subject felt that it "cared" for him/her. - Robot reminded the subject of his old dog at home – who wants to please and interact, but is slow in doing so;	- Projected identity ("Lazarus") and gender ("he") to the robot; different behaviors (tail wagging, ears, LED changes, and whole body movements) facilitated constructing a personality; - Different personalities – from really happy to vicious (protective) - Saw personality through "facial expressions" and body movement – the robot had a huge level of nonverbal communication - Saw the personality in little things the dog did - LED change, tail and ear movements and changes in speed and gaits added to the perception of the personality - Enjoyed seeing the personality of a "puppy" though tail wag, crawling, ear movements
Aibo's Emotional States	- Shaking head at seeing the pink color – excitement - Seemed to enjoy following the ball; neutral most of the time - Sense of emotion when the dog reached the flowers and rubbed its nose in them; - When the head was shaking as it saw the flowers, it looked excited, but otherwise it focused on executing the commands, not on being cute; never changed emotions or personality throughout the sessions.	- Confusion, recognition, defensive, exuberance and readiness; - Some happiness when the dog spotted the flowers, and playfulness when approaching them; - Excited when coming to the subject; - Seemed happy when it would rub its head against the bouquet of flowers and wag its tail and ears and seemed eager for attention; - Cheerful when a command is given; - Some emotions similar to real dog's emotions; - Excited when recognized flowers; on guard with the intruder; - Happy, protective, playful, obedient.

Table 58 (continued)

Subject's state	- Entertained; scared for the dog when it falls down; loved the reaction to pink color; - Independent of the dog – based on the user state before the experiment; - Excitement and curiosity at the beginning to hostility later for not performing the commands as instructed/taking too long; - Entertained - Curious, entertained when the commands were performed well; frustrated when not - Entertained; created a personality he observed in the dog by playing with it, by causing the dog to react in different ways; rooted for the robot when it could overcome obstacles the subject constructed while "playing" - Robot's not reacting with excitement at praise or during "follow me" and "come to me" commands made the subject feel less excited - Similarity between the robot and the subject's old dog endeared him to the robot even more; quite entertained	- Interested; felt cruel for not treating the robot as a real dog; emotions helped the subject recognize the robot's progress in fulfilling the commands - Curiously entertained - Entertained, but also frustrated when the robot didn't respond; - Interested to see the dog respond - Enjoyed the interaction; liked the robot having the emotions - Strongly enjoyed the interaction; got attached to the robot and named it "Bobik"; curious; positively surprised at the level of the robot's intelligence - Enjoyed the interaction; was entertained, curious, patient, longing for more behaviors/commands to interact with the robot; robot's attempts to protect the subject from the intruder were endearing

Table 58 (continued)

Subject's change of attitude	- Unrelated to the robot (participant's comment) - More hostile as the sessions carried on - More comfortable as the subject learned what the do could do right or wrong - More attached, and at the same time more frustrated and bored - More attached as the subject discovered robot's limitations and successes - Bored by the 3rd session - Grew more fond of the robot each session	- Got bored as the sessions went by - The first session was the most enjoyable – interaction seemed scripted after that - Very satisfying experience in the first sessions, but more frustrating later - Grew more involved and more interested - Interaction was more interesting in the beginning - Grew more attached from session to session - Became more protective of the robot
Why "yes" to emotions/personality in robots	- Makes the whole interaction more personal and enjoyable - Enough personality to get the job done with as little interaction as possible; as the robots become more prevalent, cannot interact without human traits/personalities - It's appealing even in machines; - That's what humans can relate to; - Would make interaction more enjoyable for certain applications - Would make it more pleasant to interact with them (only positive emotions/personality traits) - Robots with emotions would be readily accepted by most people; people would be less intimidated by robots and would be more comfortable dealing with them	- Better usability and communication of commands - Appropriate for entertainment - Makes interaction more enjoyable - Interaction would get boring without it - To make it more like interacting with real animals/people - Adds a little life to a mundane process and can get very creative - Appropriate for entertainment/pleasure purposes - Humans need emotional response - When they contribute to the overall goal for humans, e.g., to make entertainment enjoyable

Table 58 (continued)

Why "no" to emotions/personality in robots	- Emotions in robots can cause attachment in the person - Because robots are not human, but rather made by humans - For certain applications	- When it's for mundane tasks, such as bagging groceries or mopping the floor - For strictly functional tasks, such as those in factories

APPENDIX I: CODING INSTRUCTIONS FOR FREE-RESPONSE QUESTIONNAIRE GIVEN TO INDEPENDENT CODERS

At the end of an experiment, in which participants had a chance to interact with a robotic dog in a series of 4 sessions, they were asked to answer the questions below in a free-style manner:

> Please use the space below (attach additional sheets if needed) to describe your interactions with the robotic dog. Specifically, did the dog seem to have a personality? If so, what kind of personality? Also, describe any emotional states that you think the dog exhibited during your interaction. Please describe your own state during the interaction: e.g., entertained, bored, curious, etc. Did your attitude to the robotic dog change throughout the sessions? How? Finally, would you prefer robots that interact with humans to express some emotion and personality? Why?

Please read through the participants' responses and assess them according to the scales below. Not all of the participants answered every question, therefore if the answer to a scale cannot be determined from the responses, please mark it "NR" (no response). Thank you for your time!

1. On a scale from 1 to 7, please rate the level to which the participant perceived emotions and/or personality in the robot:

NR	1	2	3	4	5	6	7
	Very Low	Low	Somewhat Low	Medium	Somewhat High	High	Very High

2. On a scale from 1 to 7, please rate the level of detail in the descriptions of emotions and personality are:

NR	1	2	3	4	5	6	7
	Very Low	Low	Somewhat Low	Medium	Somewhat High	High	Very High

3. On a scale from 1 to 7, please rate the participant's boredom during the interaction:

NR	1	2	3	4	5	6	7
	Very Low	Low	Somewhat Low	Medium	Somewhat High	High	Very High

4. On a scale from 1 to 7, please rate the participant's enjoyment during the interaction:

NR	1	2	3	4	5	6	7
	Very Low	Low	Somewhat Low	Medium	Somewhat High	High	Very High

5. On a scale from 1 to 7, please rate the participant's frustration during the interaction:

NR	1	2	3	4	5	6	7
	Very Low	Low	Somewhat Low	Medium	Somewhat High	High	Very High

6. On a scale from 1 to 7, please rate the participant's contentment during the interaction:

NR	1	2	3	4	5	6	7
	Very Low	Low	Somewhat Low	Medium	Somewhat High	High	Very High

7. On a scale from 1 to 7, please rate the change in participant's attitude towards the robot throughout the sessions:

NR	1	2	3	4	5	6	7
	Overwhelmingly to the worse	Significantly to the worse	Somewhat to the worse	No change	Somewhat to the better	Significantly to the better	Overwhelmingly to the better

APPENDIX J: SAMPLE AFFECT CONFIGURATION FILE

```
<Configuration>

<Trait> // personality configuration
  <Openness>110</Openness>
  <Conscientiousness>123</Conscientiousness>
  <Extroversion>109</Extroversion>
  <Agreeableness>124</Agreeableness>
  <Neuroticism>79</Neuroticism>
</Trait>

// configuration for each emotion
<Emotion>
  <Fear>
    <TraitDependencyO>0</TraitDependencyO>
    <TraitDependencyC>0</TraitDependencyC>
    <TraitDependencyE>0</TraitDependencyE>
    <TraitDependencyA>0</TraitDependencyA>
    <TraitDependencyN>1</TraitDependencyN>
    <PriorWeight>1</PriorWeight>
    <CurrentWeight>2</CurrentWeight>
    <MinActPoint>0.5</MinActPoint>
    <MaxActPoint>7</MaxActPoint>
    <MinAmplitude>6</MinAmplitude>
    <MaxAmplitude>10</MaxAmplitude>
    <MinSlopePoint>1</MinSlopePoint>
    <MaxSlopePoint>5</MaxSlopePoint>
    <DecayRate>0.01</DecayRate>
  </Fear>

  <Disgust>
    <TraitDependencyO>0</TraitDependencyO>
    <TraitDependencyC>0</TraitDependencyC>
    <TraitDependencyE>0</TraitDependencyE>
    <TraitDependencyA>0</TraitDependencyA>
    <TraitDependencyN>0</TraitDependencyN>
    <PriorWeight>1</PriorWeight>
    <CurrentWeight>2</CurrentWeight>
    <MinActPoint>0.5</MinActPoint>
    <MaxActPoint>7</MaxActPoint>
    <MinAmplitude>6</MinAmplitude>
    <MaxAmplitude>10</MaxAmplitude>
    <MinSlopePoint>1</MinSlopePoint>
    <MaxSlopePoint>5</MaxSlopePoint>
    <DecayRate>0.01</DecayRate>
  </Disgust>
```

Personality/emotion Dependency matrix

Filtering variables

Emotion generation variables (activation point, amplitude and slope upper and lower bounds

```
<Anger>
  <TraitDependencyO>0</TraitDependencyO>
  <TraitDependencyC>0</TraitDependencyC>
  <TraitDependencyE>0</TraitDependencyE>
  <TraitDependencyA>0</TraitDependencyA>
  <TraitDependencyN>0</TraitDependencyN>
  <PriorWeight>1</PriorWeight>
  <CurrentWeight>2</CurrentWeight>
  <MinActPoint>0.5</MinActPoint>
  <MaxActPoint>7</MaxActPoint>
  <MinAmplitude>6</MinAmplitude>
  <MaxAmplitude>10</MaxAmplitude>
  <MinSlopePoint>1</MinSlopePoint>
  <MaxSlopePoint>5</MaxSlopePoint>
  <DecayRate>0.01</DecayRate>
</Anger>

<Sadness>
  <TraitDependencyO>-1</TraitDependencyO>
  <TraitDependencyC>0</TraitDependencyC>
  <TraitDependencyE>-1</TraitDependencyE>
  <TraitDependencyA>0</TraitDependencyA>
  <TraitDependencyN>0</TraitDependencyN>
  <PriorWeight>100</PriorWeight>
  <CurrentWeight>0</CurrentWeight>
  <MinActPoint>0.5</MinActPoint>
  <MaxActPoint>7</MaxActPoint>
  <MinAmplitude>6</MinAmplitude>
  <MaxAmplitude>10</MaxAmplitude>
  <MinSlopePoint>1</MinSlopePoint>
  <MaxSlopePoint>5</MaxSlopePoint>
  <DecayRate>0.01</DecayRate>
</Sadness>

<Joy>
  <TraitDependencyO>1</TraitDependencyO>
  <TraitDependencyC>0</TraitDependencyC>
  <TraitDependencyE>1</TraitDependencyE>
  <TraitDependencyA>0</TraitDependencyA>
  <TraitDependencyN>0</TraitDependencyN>
  <PriorWeight>1</PriorWeight>
  <CurrentWeight>0</CurrentWeight>
  <MinActPoint>0.5</MinActPoint>
  <MaxActPoint>7</MaxActPoint>
  <MinAmplitude>6</MinAmplitude>
  <MaxAmplitude>10</MaxAmplitude>
  <MinSlopePoint>1</MinSlopePoint>
  <MaxSlopePoint>5</MaxSlopePoint>
  <DecayRate>0.01</DecayRate>
</Joy>
```

```
  <Interest>
    <TraitDependencyO>0</TraitDependencyO>
    <TraitDependencyC>0</TraitDependencyC>
    <TraitDependencyE>0</TraitDependencyE>
    <TraitDependencyA>0</TraitDependencyA>
    <TraitDependencyN>0</TraitDependencyN>
    <PriorWeight>1</PriorWeight>
    <CurrentWeight>2</CurrentWeight>
    <MinActPoint>0.5</MinActPoint>
    <MaxActPoint>7</MaxActPoint>
    <MinAmplitude>6</MinAmplitude>
    <MaxAmplitude>10</MaxAmplitude>
    <MinSlopePoint>1</MinSlopePoint>
    <MaxSlopePoint>5</MaxSlopePoint>
    <DecayRate>0.01</DecayRate>
  </Interest>
</Emotion>

<Mood>

  <NumOfMoods> 2 </NumOfMoods>
  <RangeSD> 1 </RangeSD>  // +/- SD bounds for mood bias produced on traits

  <PositiveMood>
    <DependencyForTrait_O> 0 </DependencyForTrait_O>
    <DependencyForTrait_C> 0 </DependencyForTrait_C>      Direction of influence of
    <DependencyForTrait_E> 0 </DependencyForTrait_E>      mood on traits
    <DependencyForTrait_A> 0 </DependencyForTrait_A>      (relationship r)
    <DependencyForTrait_N> 0 </DependencyForTrait_N>
    <InfluencePercentForTrait_O> 0 </InfluencePercentForTrait_O>
    <InfluencePercentForTrait_C> 0 </InfluencePercentForTrait_C>      Relative strength of
    <InfluencePercentForTrait_E> 0 </InfluencePercentForTrait_E>      mood influence on
    <InfluencePercentForTrait_A> 0 </InfluencePercentForTrait_A>      traits
    <InfluencePercentForTrait_N> 0 </InfluencePercentForTrait_N>
  </PositiveMood>

  <NegativeMood>
    <DependencyForTrait_O> 0 </DependencyForTrait_O>
    <DependencyForTrait_C> 0 </DependencyForTrait_C>
    <DependencyForTrait_E> 0 </DependencyForTrait_E>
    <DependencyForTrait_A> 0 </DependencyForTrait_A>
    <DependencyForTrait_N> 1 </DependencyForTrait_N>
    <InfluencePercentForTrait_O> 0 </InfluencePercentForTrait_O>
    <InfluencePercentForTrait_C> 0 </InfluencePercentForTrait_C>
    <InfluencePercentForTrait_E> 0 </InfluencePercentForTrait_E>
    <InfluencePercentForTrait_A> 0 </InfluencePercentForTrait_A>
    <InfluencePercentForTrait_N> 90 </InfluencePercentForTrait_N>
  </NegativeMood>

</Mood>
```

<Attitude>

<SpecificityFactor>0.05</ SpecificityFactor> // 0.0 – most general; 0.15 – most specific
<SimilarityThreshold>0.7</ SimilarityThreshold>
<WithRevision>1</WithRevision> // 1 = revision and retention steps take place; 0 = not
<FinalCaseSelection>1</FinalCaseSelection> // 1 = highest similarity score; 2 = random roulette; 3 = highest rank

</Attitude>

</Configuration>

APPENDIX K: SAMPLE STIMULI CONFIGURATION FILE

```
<Configuration>

<NumOfStimuli>5</NumOfStimuli> // number of available percepts

<Mood>
  <ID>1</ID>
  <Name>NegativeMood</Name>
  <MappingFactor>1</MappingFactor>       // overall scaling factor
    <Stimulus0>0</Stimulus0>
    <Stimulus1>0</Stimulus1>
    <Stimulus2>1</Stimulus2>
    <Stimulus3>0</Stimulus3>
    <Stimulus4>0</Stimulus4>
```

Scaling/weighting factors for each type of percept; used for normalization purposes and to designate relative importance of each; 0 signifies “not relevant”

```
 <ID>2</ID>
 <Name>PositiveMood</Name>
<MappingFactor>1</MappingFactor>
    <Stimulus0>0</Stimulus0>
    <Stimulus1>0</Stimulus1>
    <Stimulus2>0</Stimulus2>
    <Stimulus3>-1</Stimulus3>
    <Stimulus4>0</Stimulus4>
</Mood>


<Fear>
  <MappingFactor>1</MappingFactor>       // overall scaling factor for fear
  <Stimulus0>5</Stimulus0>
  <Stimulus1>0</Stimulus1>
  <Stimulus2>21</Stimulus2>
  <Stimulus3>-7</Stimulus3>
  <Stimulus4>0</Stimulus4>
</Fear>
```

Scaling/weighting factors for each type of percept

```
<Interest>
  <MappingFactor>0</MappingFactor>
  <Stimulus0>0</Stimulus0>
  <Stimulus1>0</Stimulus1>
  <Stimulus2>0</Stimulus2>
  <Stimulus3>0</Stimulus3>
  <Stimulus4>0</Stimulus4>
</Interest>
```

```
<Disgust>
  <MappingFactor>0</MappingFactor>
  <Stimulus0>0</Stimulus0>
  <Stimulus1>0</Stimulus1>
  <Stimulus2>0</Stimulus2>
  <Stimulus3>0</Stimulus3>
  <Stimulus4>0</Stimulus4>
</Disgust>

<Anger>
  <MappingFactor>0</MappingFactor>
  <Stimulus0>0</Stimulus0>
  <Stimulus1>0</Stimulus1>
  <Stimulus2>0</Stimulus2>
  <Stimulus3>0</Stimulus3>
  <Stimulus4>0</Stimulus4>
</Anger>

<Joy>
  <MappingFactor>0.6</MappingFactor>
  <Stimulus0>0</Stimulus0>
  <Stimulus1>1</Stimulus1>
  <Stimulus2>0</Stimulus2>
  <Stimulus3>0</Stimulus3>
  <Stimulus4>0</Stimulus4>
</Joy>

<Sadness>
  <MappingFactor>7</MappingFactor>
  <Stimulus0>0</Stimulus0>
  <Stimulus1>0</Stimulus1>
  <Stimulus2>1</Stimulus2>
  <Stimulus3>0</Stimulus3>
  <Stimulus4>0</Stimulus4>
</Sadness>

<Attitudes>
<ObjectIDnumber>10<ObjectIDnumber>

</Configuration>
```

APPENDIX L: PSEUDOCODE FOR CHAPTER 5: SOFTWARE ARCHITECTURE AND IMPLEMENTATION

L.1: EMOTION UPDATE

Emotion Update function is called from the TAME Manager whenever an emotion-eliciting stimulus for any of the emotions is present.

Pseudocode:

1. Receive stimulus strength data and current trait and mood values from TAME Manager
2. Retrieve default upper and lower bounds for emotion generation variables a, g, and d; decay rate; filter weights and personality/emotion dependency matrix from Affect Configuration file
3. **for** each emotion
4. **If** stimulus strength for this emotion > 0
/*** Calculate Emotion generation Variables ***/
5. Calculate amplitude (eq. 7), slope (eq. 9) and activation point (eq. 11) for each personality trait (linear mapping, subsection 3.3.2.2.1, *Personality Influence on Base Emotion Generation*)
6. g = average amplitude across all traits (eq. 8)
7. d = average slope across all traits (eq. 10)
8. a_{trait} = average activation point across all traits (eq. 12)
9. Calculate mood-based activation point, a_{mood} (eq. 13, linear mapping, subsection 3.3.2.2.2, *Mood Influence on Base Emotion Generation*)
10. $a_{total} = (a_{trait} + a_{mood})/2$ (eq. 14)
11. Calculate base emotion value, using stimulus strength and emotion generation variables (eq. 6)
12. Update the base value with response decay (eq. 15)
13. Filter the resulting emotion value (eq. 16)
14. **End if**
15. **End for**
16. Return the updated emotion values to TAME Manager

L.2: MOOD UPDATE

Mood Update function is called from the TAME Manager whenever mood-relevant conditions are received for either Positive or Negative Mood.

Pseudocode:

1. Receive the overall condition levels and current trait values from TAME Manager
2. Retrieve personality/mood dependency and relative strength matrices, and the range within which mood can adjust personality traits (between 0 and 2 SD, with 0 indicating no effect of mood on personality, and 2 maximum effect)
3. **for** each mood

/*** current mood level is reflected in the environmental and internal conditions relevant for it ***/
4. current mood level = retrieved normalized mood conditions value
5. f**or** each personality trait
6. calculate the amount by which mood adjusts personality trait (eq. 33 and 34)
7. adjust the personality trait by that amount
8. **end for**
9. **end for**
10. pass the mood level values and updated trait values to TAME Manager

L.3: ATTITUDE UPDATE

Attitude Update function is called by the TAME Manager is called whenever an ID of an attitude-invoking object is received by the TAME Manager.

Pseudocode:

1. Retrieve default settings from Affect Configuration file: similarity threshold, specificity level, “with revision” marker, and case selection mode
2. Retrieve object ID from TAME Manager
/*** Object feature interpretation ***/
3. Based on marker interpretation table, populate object feature vector
/*** Retrieval step ***/
4. **for** each case in the case library
5. Compute similarity score and ranking (eq.26 - 28)
6. if (similarity score > similarity threshold)
7. Add the case to initial match set
8. **end if**
9. **end for**
10. **if** (case selection mode = highest similarity score)
11. retrieve the case with highest similarity score from initial match set
12. **end if**
13. **else if** (case selection mode = random roulette)
14. retrieve the case returned by random roulette function
15. **end if**
16. **else if** (case selection mode = highest ranking)
17. retrieve the case with the highest ranking from initial match set
18. **end if**
/*** Reuse Step ***/
19. Pass the best case output (attitude-based stimuli values for each emotion) to TAME Manager
/*** Revision Step ***/
20. **if** (with revision = 1)
21. retrieve user input (which is the desired output vector for the attitude-invoking object presented to the robot)
22. **if** (user input = best case output)
23. Combine object features and best case output into a new case

24. **end if**
25. **else if** (user input != best case output)
26. Combine object features and user input into a new case
/*** this object already exists in case library – update with new user input***)
27. **if** (best case object feature vector = new object feature vector)
28. Update the best case based on user input (eq. 29) and store the case in the case library
29. **end if**
/*** Learn new weights ***/
30. **for** each case in case library
31. Calculate dissimilarity score for each feature (eq. 30 and 31)
32. **end for**
33. Update the weight for the feature with the lowest dissimilarity score (eq. 32)
34. Update the feature weight table with the new weight
35. **end if**
36. **end if**
37. Add the new case to the case library

L.4 TAME MANAGER

TAME Manager acts as a central control point, invoking emotion, attitude and mood updates whenever the corresponding perceptual information becomes available and relaying TAME variables to the Robot Executable via TAME Communication Manager.

Pseudocode:

1. Connect to TAME Communications Manager
2. **Repeat**
3. Retrieve incoming data from TAME Communications Manager
4. **If** incoming data contains emotion-eliciting stimulus
5. Call update emotion function
6. Return updated emotion values to TAME Communications Manager
7. **End if**
8. **If** internal or external conditions for mood changed
9. Call update mood function
10. Return updated mood values to TAME Communications Manager
11. **End if**
12. **If** attitude-invoking object is present
13. Call update attitude function
14. Call update emotion function, for attitude-based emotion generation
15. Return updated emotion values to TAME Communications Manager
16. **End if**
17. **Until** (disconnected)

L.5: TAME COMMUNICATIONS MANAGER

TAME Communications Manager ensures proper communication between Robot Executable, TAME Manager and Stimuli Interpreter.

Pseudocode:

1. Connect to Robot Executable
2. Connect to TAME Manager
3. Connect to Stimuli Interpreter
4. **Repeat**:
5. Retrieve sensor data from Robot Executable
6. Pass the sensor data to Stimuli Interpreter
7. Retrieve processed perceptual info (stimuli strengths for emotion, overall condition influence for mood, and object ID marker for attitude generation) from Stimuli Interpreter
8. Pass the processed perceptual info to TAME Manager
9. Retrieve updated TAME variables (emotion, mood and trait values) from TAME Manager
10. Pass the TAME variables to the Robot Executable
11. **Until** (disconnected)

L.6: STIMULI INTERPRETER

Stimuli Interpreter processes sensor data received from Robot Executable via TAME Communication Manager, and relays the processed data to TAME Manager.

Pseudocode:

1. Connect to TAME Communications Manager
2. Retrieve default settings for perceptual processing from Stimuli Configuration file: scaling/weighting factors for emotion- and mood-relevant percepts (where 0 signifies irrelevance); number of percepts; mapping factors for each mood and emotion
3. **repeat:**
4. Retrieve sensor data from TAME Communication Manager
5. **for** each emotion
6. $s_i = 0$
7. N = 0
8. **for** each percept
   /*** calculate stimulus strength, eq. ***/
9. **if** (percept is relevant to emotion *i*)
10. $s_i = s_i$ + percept * scaling_factor
11. N = N + 1
12. **end if**
13. **end for**
14. **if** N >0

```
                s_i = s_i
        end if
pass stimulus strength s_i  for emotion i to TAME Communications Manager
end for
for each mood
        m_i = 0
        N = 0
        for each percept
        /*** calculate mood condition level, eq. ; ; both environmental conditions (e.g.,
        lighting or noise level) and internal conditions (e.g., battery level or internal
        temperature) are included  ***/
                if (percept is relevant to mood i)
                        m_i = m_i + percept * scaling_factor
                        N = N + 1
                end if
        end for
        if N >0
                m_i = m_i
        end if
pass overall conditions average m_i for mood i to TAME Communications Manager
end for
for each percept
        /*** pass object ID for attitudes ***/
        if (percept is relevant for attitudes)
                pass the object data to TAME Communications Manager
        end if
end for
until (disconnected)
```

APPENDIX M: SNAPSHOTS OF INTERFACE FOR SOLICITING USER INPUT DURING THE REVISION STAGE OF ATTITUDE CBR PROCESS

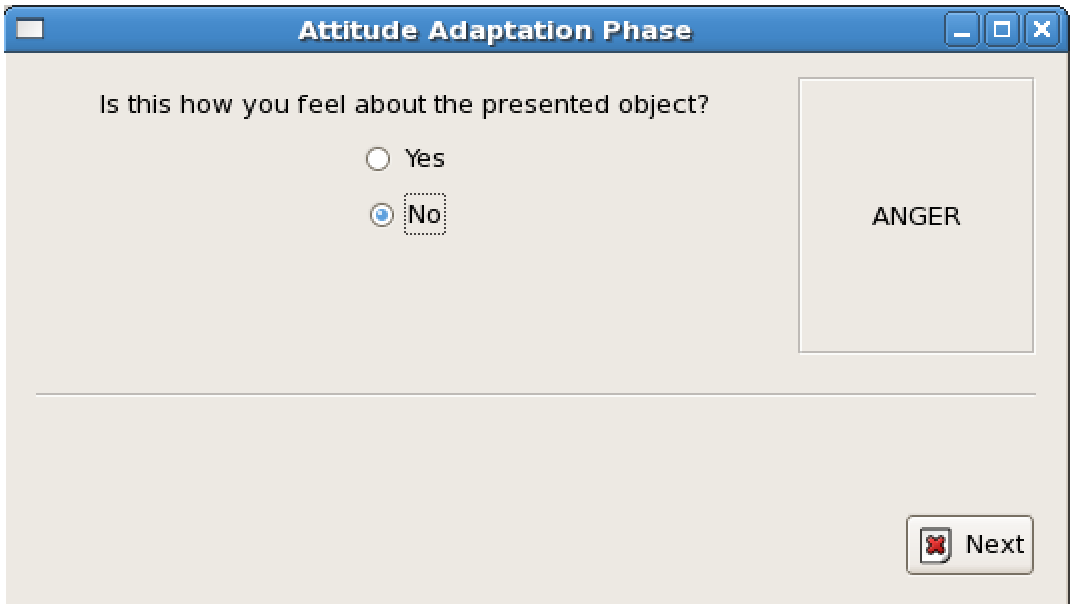


Figure 102: First screen: the user disagrees with the attitude displayed by the robot

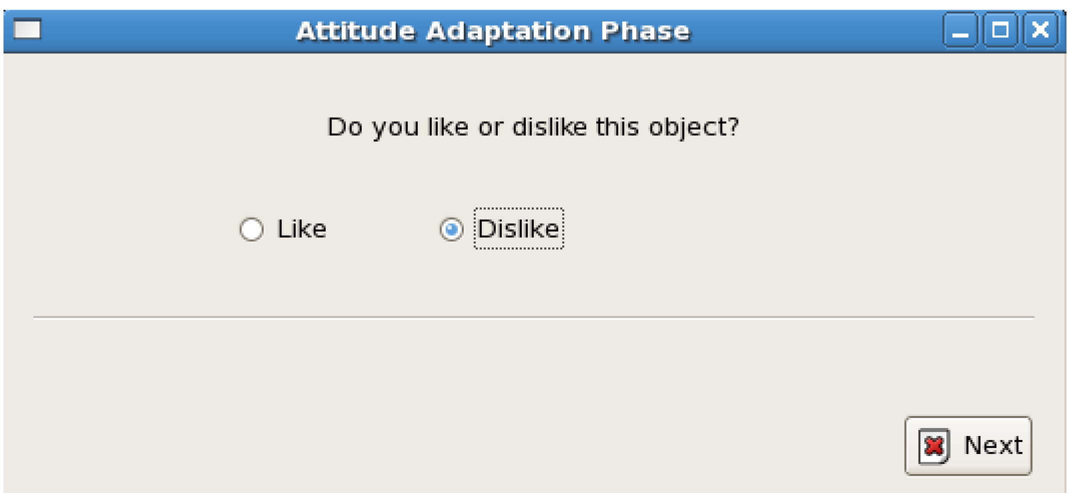


Figure 103: The second screen: the user dislikes the presented object

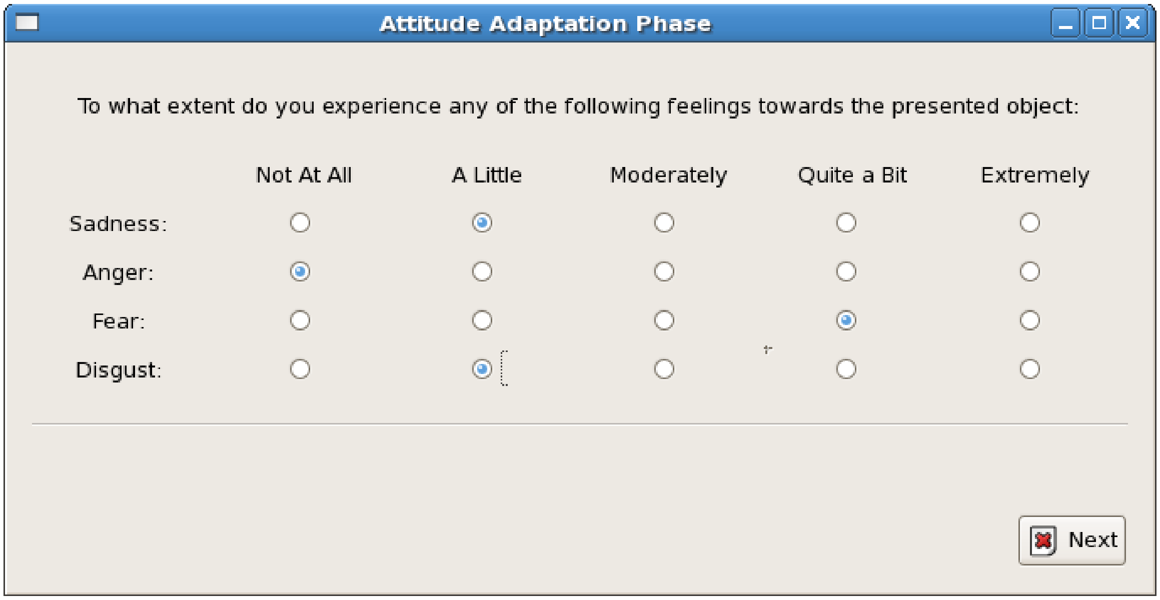


Figure 104: Screen 3: the user selects the desired attitudes for the presented object: Fear is dominant

APPENDIX N: SCREENSHOTS OF THE ONLINE SURVEY ON ROBOT BEHAVIOR AS PRESENTED TO PARTICIPANTS

Survey on Robot Behavior

Dear Participant:

You are invited to take a quick fun survey, in which we will ask your opinions on how you think the robot in a few short videos behaves. This survey is a pilot in a larger study on robot behavior, and will give us some valuable information on further design choices. Thank you very much for your interest and your willingness to participate in this survey - your sincere opinions will be greatly appreciated.

For this survey, you will be asked to watch 6 short video clips (< 1 minute each) in which a humanoid robot will perform some tasks. After each video you will be asked to briefly state your opinion on what you saw, as well as answer a number of multiple-choice questions. At the end of the survey, you will be asked to complete a short demographics questionnaire. The time it takes to complete the survey should be approximately 10-20 minutes. If you cannot complete the survey in one seating, simply close the browser and come back at another time by clicking the link provided in the e-mail you received. Your answers up to that point will be saved, and, provided you use the same computer, you will be taken to the same point where you left off.

If you have any further questions or concerns, please don't hesitate to contact Lilia Moshkina, a survey administrator, by e-mail at lilia@cc.gatech.edu or by mail at

College of Computing
Georgia Institute of Technology
Atlanta, GA 30332-0280

Click to Next Page

5%

Survey on Robot Behavior

1. Title of Research Project
Survey on Robot Behavior - Pilot Study

2. Principal Investigator
Ronald C. Arkin

3. Purpose of Research
You are being asked to volunteer to complete a survey asking your opinions on robot behavior in 6 short video clips. This information will help us make further design decisions for a larger robot study, and will eventually help achieve a goal of making robots more user-friendly.

4. Procedure
You are asked to complete a survey asking your opinions on robot behavior in 6 short video clips. This survey will be fully web-based, and no paper or direct interaction with others will be required. Your time commitment should not exceed 20 minutes.

5. Foreseeable risks or discomforts
This research involves minimal risk, similar to that involved in any everyday activity, for example, reading your e-mail.

6. Benefits
Although there are no direct benefits to you, your opinions may help design robots that are user-friendly and comfortable to be around.

7. Compensation/Costs
Neither costs nor compensation are involved with this survey.

8. Confidentiality
The following procedures will be followed to keep your personal information confidential in this study: The data that is collected about you will be kept private to the extent allowed by law. To protect your privacy, your records will be kept under a code number rather than by name, and stored securely under lock and key; only study staff will be allowed to look at them. Your name and any other fact that might point to you will not appear when results of this study are presented or published. You should be aware, however, that the survey is not being run from a "secure" https server of the kind typically used to handle credit card transactions, so there is a small possibility that responses could be viewed by unauthorized third parties (e.g., computer hackers). Also, in general the web page software will log the IP address of the machine you use to access this page (e.g., 102.403.506.807), but otherwise no other information will be stored unless you explicitly enter it.

To make sure that this research is being carried out in the proper way, the Georgia Institute of Technology IRB will review study records. Your privacy will be protected to the extent allowed by law.

9. Injury/Adverse Reactions
Reports of injury or reaction should be made to Ronald C. Arkin at (404) 894-8209. Neither the Georgia Institute of Technology nor the principal investigator has made provision for payment of costs associated with any injury resulting from participation in this study.

10. Contact Person
If you have questions about the research, call or write to the principal investigator at:

Ronald C. Arkin
College of Computing
Georgia Institute of Technology
Atlanta, GA 30332-0280
(404) 894-8209 (voice)
(404) 385-5251 (fax)
arkin@cc.gatech.edu

11. Voluntary Participation/Withdrawal
You have rights as a research volunteer. Taking part in this study is completely voluntary. If you do not take part, there is no penalty. You may stop taking part in this study at any time with no penalty. If you have any questions about your rights as a research volunteer, call or write:

11. Voluntary Participation/Withdrawal

You have rights as a research volunteer. Taking part in this study is completely voluntary. If you do not take part, there is no penalty. You may stop taking part in this study at any time with no penalty. If you have any questions about your rights as a research volunteer, call or write:

The Institutional Review Board,
Office of Research Compliance
505 Tenth Street, 3rd Floor
Georgia Institute of Technology
Atlanta, Georgia 30332-0420
(404) 894-6942 (voice)
(404) 894-2081 (fax)
IRB@gatech.edu

Thank you for your participation.

By selecting 'I agree', I certify that I am 18 years old or older, have read the information in this consent form, and agree to participate in this survey. *

- I agree
- I do not agree

APPROVED *Consent Form Approved by Georgia Tech IRB: March 12, 2010 - Indefinite*

Click to Go Back | Click to Next Page

11%

Survey on Robot Behavior

Please play and watch the video clip below. Pay attention to the ***MANNER*** in which the robot in the clip behaves, as you will be asked a few questions about it. Feel free to watch the video as many times as you need to answer the questions; you can always come back to this page by pressing "Click to Go Back" button.

CLIP 1: FEAR

1. In your own words, please briefly describe your opinion on the *MANNER* in which the robot behaves in this video clip, for example, "neutral", "reserved", "excited", "clumsy", etc.

Click to Go Back | Click to Next Page

17%

Survey on Robot Behavior

2. In your opinion, which emotion, if any, from the list below best describes the robot's behavior in this clip?

- Interest
- Sadness
- Joy
- Fear
- Disgust
- Anger
- None
- Other - please specify

3. If you selected an emotion from the list above, to what extent was this emotion expressed by the robot in the clip?

A Little | Moderately | Quite a Bit | Extremely | N/A

Click to Go Back | Click to Next Page

23%

Survey on Robot Behavior

Please play and watch the video clip below. Pay attention to the ***MANNER*** in which the robot in the clip behaves, as you will be asked a few questions about it. Feel free to watch the video as many times as you need to answer the questions; you can always come back to this page by pressing "Click to Go Back" button.

CLIP 2: JOY

1. In your own words, please briefly describe your opinion on the *MANNER* in which the robot behaves in this video clip, for example, "neutral", "reserved", "excited", "clumsy", etc.

Click to Go Back | Click to Next Page

17%

Survey on Robot Behavior

2. In your opinion, which emotion, if any, from the list below best describes the robot's behavior in this clip?

- Interest
- Sadness
- Joy
- Fear
- Disgust
- Anger
- None
- Other - please specify

3. If you selected an emotion from the list above, to what extent was this emotion expressed by the robot in the clip?

A Little | Moderately | Quite a Bit | Extremely | N/A

Click to Go Back | Click to Next Page

23%

Survey on Robot Behavior

Please play and watch the video clip below. Pay attention to the ***MANNER*** in which the robot in the clip behaves, as you will be asked a few questions about it. Feel free to watch the video as many times as you need to answer the questions; you can always come back to this page by pressing "Click to Go Back" button.

CLIP 3: EXTRAVERSION

1. In your own words, please briefly describe your opinion on the *MANNER* in which the robot behaves in this video clip, for example, "neutral", "reserved", "excited", "clumsy", etc.

Click to Go Back | Click to Next Page

17%

APPENDIX N

Survey on Robot Behavior

8. Please use this list of common human traits to describe the robot from the clip you just saw as accurately as possible. For each trait, please select how accurately that trait describes the robot.

	Extremely Inaccurate	Very Inaccurate	Moderately Inaccurate	Slightly Inaccurate	Neither Accurate Nor Inaccurate	Slightly Accurate	Moderately Accurate	Very Accurate	Extrem Accur
Withdrawn									
Energetic									
Talkative									
Extraverted									
Quiet									
Bashful									
Shy									
Bold									

APPENDIX N

Below are definitions for your information:

Bashful: socially shy or timid; diffident; self-conscious.
Bold: fearless before danger; intrepid; impudent; presumptuous.
Energetic: operating with or marked by vigor or effect; vigorous.
Extraverted: gregarious and unreserved; outgoing.
Quiet: restrained in style; understated; making little or no noise.
Shy: marked by reserve or diffidence; reserved; wary.
Talkative: full of trivial conversation; loquacious; garrulous; voluble.
Withdrawn: not friendly or sociable; aloof; detached; emotionally unresponsive.

Click to Go Back   Click to Next Page

47%

Survey on Robot Behavior

Please play and watch the video clip below. Pay attention to the ***MANNER*** in which the robot in the clip behaves, as you will be asked a few questions about it. Feel free to watch the video as many times as you need to answer the questions; you can always come back to this page by pressing "Click to Go Back" button.

CLIP 4: INTROVERSION

1. In your own words, please briefly describe your opinion on the *MANNER* in which the robot behaves in this video clip, for example, "neutral", "reserved", "excited", "clumsy", etc.

Click to Go Back | Click to Next Page

17%

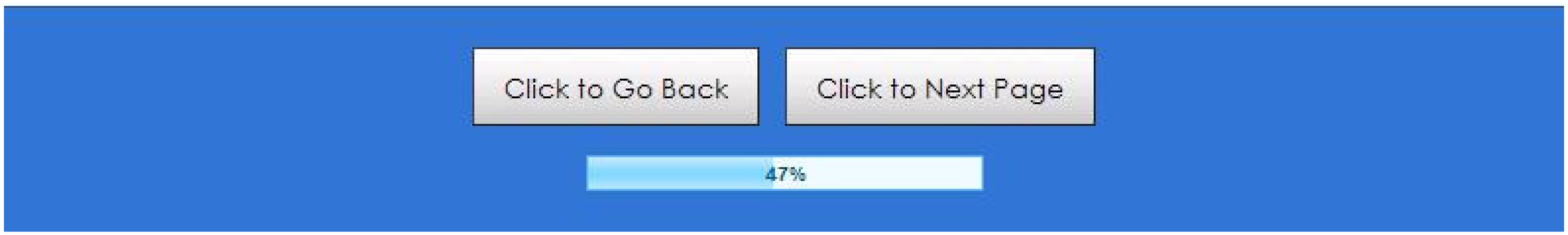
Survey on Robot Behavior

8. Please use this list of common human traits to describe the robot from the clip you just saw as accurately as possible. For each trait, please select how accurately that trait describes the robot.

	Extremely Inaccurate	Very Inaccurate	Moderately Inaccurate	Slightly Inaccurate	Neither Accurate Nor Inaccurate	Slightly Accurate	Moderately Accurate	Very Accurate	Extrem Accur
Withdrawn									
Energetic									
Talkative									
Extraverted									
Quiet									
Bashful									
Shy									
Bold									

Below are definitions for your information:

Bashful: socially shy or timid; diffident; self-conscious.
Bold: fearless before danger; intrepid; impudent; presumptuous.
Energetic: operating with or marked by vigor or effect; vigorous.
Extraverted: gregarious and unreserved; outgoing.
Quiet: restrained in style; understated; making little or no noise.
Shy: marked by reserve or diffidence; reserved; wary.
Talkative: full of trivial conversation; loquacious; garrulous; voluble.
Withdrawn: not friendly or sociable; aloof; detached; emotionally unresponsive.



Survey on Robot Behavior

Please play and watch the video clip below. Pay attention to the ***MANNER*** in which the robot in the clip behaves, as you will be asked a few questions about it. Feel free to watch the video as many times as you need to answer the questions; you can always come back to this page by pressing "Click to Go Back" button.

CLIP 5: POSITIVE MOOD

1. In your own words, please briefly describe your opinion on the *MANNER* in which the robot behaves in this video clip, for example, "neutral", "reserved", "excited", "clumsy", etc.

Click to Go Back | Click to Next Page

17%

Survey on Robot Behavior

12. Please select the answer that best describes the extent to which you think the robot in the clip may be experiencing each of the feelings or emotions below

	Very Slightly or Not At All	A Little	Moderately	Quite a Bit	Extremely
Distressed					
Enthusiastic					
Irritable					
Active					
Upset					
Determined					
Interested					
Happy					
Nervous					
Excited					
Afraid					
Depressed					

Click to Go Back   Click to Next Page

70%

Survey on Robot Behavior

Please play and watch the video clip below. Pay attention to the ***MANNER*** in which the robot in the clip behaves, as you will be asked a few questions about it. Feel free to watch the video as many times as you need to answer the questions; you can always come back to this page by pressing "Click to Go Back" button.

CLIP 6: NEGATIVE MOOD

1. In your own words, please briefly describe your opinion on the *MANNER* in which the robot behaves in this video clip, for example, "neutral", "reserved", "excited", "clumsy", etc.

Click to Go Back Click to Next Page

17%

Survey on Robot Behavior

12. Please select the answer that best describes the extent to which you think the robot in the clip may be experiencing each of the feelings or emotions below

	Very Slightly or Not At All	A Little	Moderately	Quite a Bit	Extremely
Distressed					
Enthusiastic					
Irritable					
Active					
Upset					
Determined					
Interested					
Happy					
Nervous					
Excited					
Afraid					
Depressed					

Click to Go Back   Click to Next Page

70%

Survey on Robot Behavior

15. What is your gender?

- Male
- Female

16. What is your age?

-- Please Select --

17. What is the highest level of education you have achieved?

- High School
- Bachelor's Degree
- Postgraduate Degree

18. If you have completed your Bachelor's or postgraduate degree, what was your major?

19. What is your current occupation?

Click to Go Back | Click to Next Page

88%

Survey on Robot Behavior

20. Have you had any robotics experience?

- ◯ Yes
- ◯ No
- ◯ Not sure

21. How much practical experience have you had, if any, with EACH of the following technologies?

	Very Significant	Significant	Some	A Little	None
Personal Computers	◯	◯	◯	◯	◯
Internet	◯	◯	◯	◯	◯
Robots	◯	◯	◯	◯	◯
Video Games	◯	◯	◯	◯	◯

22. If you had significant previous personal, work, or education-related experience with robots, what types of robots were they?

- ☐ Industrial Robots
- ☐ Research Robots
- ☐ Military Robots
- ☐ Entertainment Robots
- ☐ Service Robots
- ☐ Humanoid Robots
- ☐ Other - please specify ____________

23. And, finally, please help us figure out if this survey worked. Did you have any difficulty answering the questions? If so, why?

Click to Go Back

Finished? Submit your Survey

97%

APPENDIX O: CONSENT FORM FOR SEARCH-AND-RESCUE EXPERIMENT

GEORGIA INSTITUTE OF TECHNOLOGY
CONSENT TO BE A RESEARCH PARTICIPANT

Project Title: Human-Robot Interaction Study I

Investigators: *Arkin, R.C., Ph.D., Moshkina, L., and Park, S.*

Protocol and Consent Title: *Human-Robot Interaction Study I consent to be a research participant; Protocol # (to be assigned); Date: April 16, 2010.*

You are being asked to be a volunteer in a research study.

Purpose:
The purpose of this study is to evaluate robotic behaviors in a human-robot interaction task. The results will help improve the design of robotic behaviors and promote robotics research. We expect to enroll 45-60 people in this study.

Exclusion/Inclusion Criteria:
Participants in this study must be over 18 years old.

Procedures:
You will be randomly assigned to one of three conditions. If you decide to be in the study, your participation will be limited to a single session which should not exceed 30 minutes. You will be invited to GA Tech Mobile Robot Laboratory, where you will first complete an initial questionnaire, then interact with a small humanoid robot, and finally complete 3 more questionnaires and a short interview. Your interaction with the robot will last 3-5 minutes, and will take place in a mock-up search-and-rescue laboratory setting. With your permission, we will videotape your interaction with the robot and your interview with a research assistant. The post-study questionnaires will ask you about your impressions of the robot, your current mood state, and your demographics. During the interview, you will be asked additional questions about your interaction with the robot. You may stop at any time and for any reason.

Risks or Discomforts:
The risks involved are no greater than those involved in playing a video or role-playing game.

Benefits:
You are not likely to benefit in any way from joining this study, however, you will have an opportunity to interact with a humanoid robot and expand your knowledge about robots in general.

Compensation to You:
You will receive 1/2 Experimetrix credits for 30 minutes of participation.

Confidentiality:
The following procedures will be followed to keep your personal information confidential in this study: The data collected about you will be kept private to the extent allowed by law. To protect your privacy, your records will be kept under a code number rather than by name. Your records, including videotapes, will be kept in locked files and only study staff will be allowed to look at them; the videotapes will transcribed and destroyed after the full data analysis is compete. Your name and any other fact that might point to you will not appear when results of this study are presented or published; we will ask your written permission for use of any videos or still photographs in demos and publications. Your privacy will be protected to the extent allowed by law. To make sure that this research is being carried out in the proper way, the Georgia Institute of Technology IRB may review study records. The Office of Human Research Protections may also look over study records during required reviews.

Costs to You:
There are no costs to you, other than your time, for being in this study.

In Case of Injury/Harm:
If you are injured as a result of being in this study, please contact Principal Investigator, Ronald C. Arkin, Ph.D., at telephone (404) 894- 8209. Neither the Principal Investigator nor Georgia Institute of Technology has made provision for payment of costs associated with any injury resulting from participation in this study.

Participant Rights:

- Your participation in this study is voluntary. You do not have to be in this study if you don't want to be.
- You have the right to change your mind and leave the study at any time without giving any reason and without penalty.
- Any new information that may make you change your mind about being in this study will be given to you.

- You will be given a copy of this consent form to keep.
- You do not waive any of your legal rights by signing this consent form.

Conflict of Interest:
None.

Questions about the Study:
If you have any questions about the study, you may contact Dr. Ronald C. Arkin at telephone (404) 894- 8209 or at arkin@cc.gatech.edu.

Questions about Your Rights as a Research Participant:

If you have any questions about your rights as a research participant, you may contact

Ms. Melanie Clark, Georgia Institute of Technology
Office of Research Compliance, at (404) 894-6942.

If you sign below, it means that you have read the information given in this consent form, and you would like to be a volunteer in this study.

____
Participant Name (printed)

______________________________________ ____________
Participant Signature Date

______________________________________ ____________
Signature of Person Obtaining Consent Date

APPENDIX P: VIDEO RELEASE FORM FOR NAO EXPERIMENTS

Mobile Robot Laboratory Video Release Form

Georgia Tech Mobile Robot Laboratory may use the video footage or still photographs from the videos showing my appearance for their research, including displaying at conferences, briefings, workshops, etc. This footage or still photographs will be used only for research or educational purposes.

By signing below, I am indicating that I accept the stipulations of my releasing my video as stated above.

Participant Name (Printed) ____
Participant Signature ____
Administrator Signature ____
Date ____

APPENDIX Q: EXPERIMENTER SCRIPT FOR SEARCH-AND-RESCUE EXPERIMENT

Greet the subject in TSRB lobby:

E: Hi, my name is Lilia, and you must be (name). Thank you for coming today. Before we start the experiment, please read over and sign this consent form, giving us permission to have you as a participant.

The subject fills out the consent form. Answer any questions he/she may have.

E: I also need your permission to use portions of the video we will be recording. I won't show your face, so it would be unlikely that anyone could recognize you. Would you please sign this video release?

The subject signs the release.

E: Great! First, I need you to fill this brief questionnaire. Please fill out both sides.

Give "PANAS-person" questionnaire to the subject. The subject fills out the questionnaire.

E: Thank you! Let's go down to the lab.

Take the subject to the lab and stopped in front of the back door.

E: Please wait here for a moment; I'll be back with you once I start the video recording.

Start the recording and the robot; go back to get the subject and lead him/her through the door. Walk into the lab behind the subject.

E: Now, for this experiment, you will be a search-and-rescue site inspector. A recent accidental explosion caused a lot of damage to this portion of the lab, and although initial efforts have been undertaken, the work is still in progress. Take a few moments to kind of take it all in.

The subject takes a look around.

E: Done? Now, please stand at the first red cross marker on the floor next to the platform. During the next couple of minutes, our Search and Rescue robot will be your guide to the site. Remember, this robot has sensors that can perceive properties of the environment that people don't have abilities to sense. All right, please do not touch the robot, and save all your questions for later. Now please proceed to the next cross marker at the end of the platform and turn around to face the robot. Now can begin.

Start cfgedit**.**

Go into the cubicle, and watch the robot. Turn off the lights appropriately, and rescue the robot if needed. Once the experiment is over (when the participant is either beyond the cross mark, or the robot has stopped moving for 3 seconds), come out and take the participant to the questionnaire table.

E: The interaction part of the experiment is over. Please fill out these questionnaires, on both sides. The first one is just like the one you already filled out, and the second one is similar, but about the robot instead, closed to the end of the interaction. Fill out everything you have on the board and let me know when you are done. I'll then ask you a few short questions, and the experiment will be finished.

Turn the camera and the robot off while the participant is filling out the questionnaires. When he/she is done, move the camera over to the participant and start recording the interview. The interview is semi-structured, so be flexible: if they already mentioned something that is asked in the next question, rephrase it slightly not to appear repetitive; ask additional questions if you hear something unusual or unexpected.

E: Done? All right, now I'll ask you a few brief questions. I will be recording our conversation, but not the faces, just the sound.

Once the interview is over, turn off the camera, then compensate the subject (either with a gift card, in which case a compensation form will need to be filled out, or through Experimetrix), and lead the subject back to the lobby.

E: Thank you very much for your participation. Here is your gift card – please fill out this compensation form for me (or "You will receive your Experimetrix credit later today"). Let me know if you have any questions; otherwise, let me take you back to the lobby.

APPENDIX R: NAO'S SPEECH DURING THE SEARCH-AND-RESCUE TOUR

Hello, Inspector. Thank you for coming today. For the next few minutes, I will be your guide on this accident site. Let me show you around.

The accident happened a few days ago, and the site has been mainly secured, with certain safety precautions in place.

As you can see, there is still a lot of work to be done.

That was unexpected.

Let's proceed nonetheless.

The debris you see up ahead is a result of a recent explosion. It has not been taken care of yet.

Something is wrong!

Inspector, the structural integrity of this site has been compromised, and we need to evacuate immediately.

Please proceed to the exit.

APPENDIX S: PANAS-T WITH REGARDS TO NAO ROBOT FOR SEARCH-AND-RESCUE EXPERIMENT

ROBOT – DURING THE SECOND HALF

Circle the answer that best describes the extent you think the robot may have been experiencing each of the feelings or emotions below, *during THE SECOND HALF OF THE INTERACTION.*

Interested	*very slightly or not at all*	*a little*	*moderately*	*quite a bit*	*extremely*
Distressed	*very slightly or not at all*	*a little*	*moderately*	*quite a bit*	*extremely*
Excited	*very slightly or not at all*	*a little*	*moderately*	*quite a bit*	*extremely*
Upset	*very slightly or not at all*	*a little*	*moderately*	*quite a bit*	*extremely*
Strong	*very slightly or not at all*	*a little*	*moderately*	*quite a bit*	*extremely*
Guilty	*very slightly or not at all*	*a little*	*moderately*	*quite a bit*	*extremely*
Scared	*very slightly or not at all*	*a little*	*moderately*	*quite a bit*	*extremely*
Hostile	*very slightly or not at all*	*a little*	*moderately*	*quite a bit*	*extremely*
Depressed	*very slightly or not at all*	*a little*	*moderately*	*quite a bit*	*extremely*
Enthusiastic	*very slightly or not at all*	*a little*	*moderately*	*quite a bit*	*extremely*
Proud	*very slightly or not at all*	*a little*	*moderately*	*quite a bit*	*extremely*
Irritable	*very slightly or not at all*	*a little*	*moderately*	*quite a bit*	*extremely*
Alert	*very slightly or not at all*	*a little*	*moderately*	*quite a bit*	*extremely*

Participant #

TURN OVER →

Ashamed	*very slightly or not at all*	*a little*	*moderately*	*quite a bit*	*extremely*
Inspired	*very slightly or not at all*	*a little*	*moderately*	*quite a bit*	*extremely*
Happy	*very slightly or not at all*	*a little*	*moderately*	*quite a bit*	*extremely*
Determined	*very slightly or not at all*	*a little*	*moderately*	*quite a bit*	*extremely*
Attentive	*very slightly or not at all*	*a little*	*moderately*	*quite a bit*	*extremely*
Jittery	*very slightly or not at all*	*a little*	*moderately*	*quite a bit*	*extremely*
Nervous	*very slightly or not at all*	*a little*	*moderately*	*quite a bit*	*extremely*
Active	*very slightly or not at all*	*a little*	*moderately*	*quite a bit*	*extremely*
Afraid	*very slightly or not at all*	*a little*	*moderately*	*quite a bit*	*extremely*

APPENDIX T: POST QUESTIONNAIRE FOR SEARCH-AND-RESCUE EXPERIMENT

Please reflect back on your interaction with the robot as you consider the questions below.

For the next 3 questions, please rate your impressions of the robot by circling the most appropriate number on the scale:

1. In your opinion, the robot's <u>BEHAVIOR</u> was:

Confusing				Clear
1	2	3	4	5

Unreadable				Easy to Read
1	2	3	4	5

Inconsistent				Consistent
1	2	3	4	5

Hard to Understand				Easy to Understand
1	2	3	4	5

Inexpressive				Expressive
1	2	3	4	5

2. In your opinion, the robot's <u>REQUEST TO LEAVE</u> was:

Ignorable				Compelling
1	2	3	4	5

Inappropriate				Appropriate
1	2	3	4	5

Ineffective				Persuasive
1	2	3	4	5

Insincere				Sincere
1	2	3	4	5

Unconvincing				Convincing
1	2	3	4	5

3. In your opinion, the robot APPEARED:

Fake				Natural
1	2	3	4	5

Machinelike				Humanlike
1	2	3	4	5

Unconscious				Conscious
1	2	3	4	5

Artificial				Lifelike
1	2	3	4	5

Inert				Interactive
1	2	3	4	5

APPENDIX T

Participant #

4. During your interaction with the robot, did the ROBOT express any of the EMOTIONS below? Please circle the number describing the degree to which any of the emotions were expressed.

ANGER

Not at all	A Little	Moderately	Quite a Bit	Extremely
1	2	3	4	5

JOY

Not at all	A Little	Moderately	Quite a Bit	Extremely
1	2	3	4	5

FEAR

Not at all	A Little	Moderately	Quite a Bit	Extremely
1	2	3	4	5

DISGUST

Not at all	A Little	Moderately	Quite a Bit	Extremely
1	2	3	4	5

SADNESS

Not at all	A Little	Moderately	Quite a Bit	Extremely
1	2	3	4	5

INTEREST

Not at all	A Little	Moderately	Quite a Bit	Extremely
1	2	3	4	5

5. To what extent, if any, did the robot's behavior, and not its words, influenced YOUR DECISION to comply with the request to leave?

Not at all	A Little	Moderately	Quite a Bit	Extremely
1	2	3	4	5

6. In your own words, please briefly describe your interaction with the robot. Please mention if there were any changes in your thoughts or feelings throughout the interaction.

APPENDIX U: DEMOGRAPHICS QUESTIONNAIRE FOR SEARCH-AND-RESCUE STUDY

Demographics Questionnaire

1. What is your gender?
   a.) Female
   b.) Male

2. What is your age?
   a.) Under 20 years old
   b.) Between 20 and 30
   c.) Between 30 and 40
   d.) Between 40 and 50
   e.) 50 or older

3. What is the highest level of education you've achieved? What was your major?
   a.) High School
   b.) Currently working on my Bachelor's ________________
   c.) Bachelor's ________________
   c.) Master's ________________
   d.) Currently working on my Ph.D. ________________
   e.) Ph.D. ________________
   f.) Other ____________________

4. Do you describe yourself as technical (having extensive experience or interest in a technical field, such as engineering, computing, math, etc)?

   a.) Yes
   b.) Somewhat
   c.) No

8. What is your level of computer experience?

   a.) None: Never used a computer before
   b.) Limited: Occasionally use a computer for tasks like e-mail, web browsing or word processing
   c.) User Level: Regularly use a computer for tasks like e-mail, web browsing or word processing
   d.) Advanced User: Have downloaded and installed at least one program from the Internet
   e.) Programmer Level: Some programming language or network administration experience
   f.) Advanced Programmer: Extensive training or experience in multiple programming languages

9. Have you ever interacted with robots? Please circle all that apply.

   a.) Never
   b.) Very limited interaction
   c.) Interaction experience with military robots
   b.) Interaction experience with industrial robots
   c.) Interaction experience with mobile robots
   d.) Interaction experience with entertainment/educational robots
   e.) Interaction experience with humanoid robots
   f.) Other – please specify ____

APPENDIX V: SEMI-STRUCTURED INTERVIEW FOR SEARCH-AND-RESCUE STUDY

Theme 1: Assessing understanding of the situation and robot's behavior

1. What could you tell about the situation from the way the robot behaved?
2. How did the situation change, in your understanding?
3. In what way was the robot reacting to the changes in conditions, if at all?

If previous questions are not answered fully:

a. Was it showing any reaction?
b. Did it behave differently after the lights went off?

4. What do you think was happening when the robot said "Something is wrong."?
5. How dangerous would you say the site was at that point?
6. Given that the situation became dangerous, how appropriate was the way the robot behaved and reacted? Why?

Theme 2: Assessing robot's effect on participant's subjective feelings

1. How did the way the robot was acting make you feel?
2. How did your feelings change during the interaction, if at all?

If previous questions are not answered fully:

a. What did you feel after the robot asked you to leave?)
b. Were you anxious to go? Were you calm and reluctant?
c. Did it make you more anxious and in a hurry to leave, or did it make you pretty calm?

3. How did the way in which the robot behaved Influence your decision to leave, if at all?

Theme 3: Assessing robot's expressive behavior

1. What was the robot expressing during the interaction?
2. How did the expressions change, if at all, from the beginning to the end?

If the previous questions are not answered fully:

a) How emotional was the robot?
b) Did you notice any nervousness or anxiety in its behavior or voice? What in particular?
c) Did you see if it became fearful at any point during the interaction? When?

Theme 4: Assessing usefulness of expressive behavior in general

1. Let's suppose that robots become very common in our everyday life. How important would it be for them to be expressive?
2. What advantages do you think expressive robots may have?

If the previous questions were not answered fully:

a. Do you think robot's expressions could help people understand what's going on, both around them and in robot's mind?

APPENDIX W: FLYER FOR SEARCH-AND-RESCUE STUDY

Georgia Tech College of Computing
Defining the New Face of Computing

STUDY PARTICIPANTS WANTED!!!

The Georgia Tech Mobile Robot Lab is conducting a human-robot interaction study, and **YOU** have a chance to interact with a **HUMANOID ROBOT** Nao! The study will take place in Tech Square (TSRB building), in an exciting mock-up search-and-rescue setting, and would require no more than ***30 MIN*** of your time. No Computer Science or Robotics background is required, but you have to be over 18 years old. We encourage people of all backgrounds and education levels to participate, but prefer those without robotics experience.

As a token of our appreciation, you will receive a **$10 GIFT CARD FROM STARBUCKS.**

If you are interested in participating, please contact Lilia Moshkina at **lilia@cc.gatech.edu** or SungHyun Park at **gte246z@gatech.edu.**

GA Tech Mobile Robot Lab Study
gte246z@gatech.edu
lilia@cc.gatech.edu

GA Tech Mobile Robot Lab Study
gte246z@gatech.edu
lilia@cc.gatech.edu

GA Tech Mobile Robot Lab Study
gte246z@gatech.edu
lilia@cc.gatech.edu

GA Tech Mobile Robot Lab Study
gte246z@gatech.edu
lilia@cc.gatech.edu

GA Tech Mobile Robot Lab Study
gte246z@gatech.edu
lilia@cc.gatech.edu

GA Tech Mobile Robot Lab Study
gte246z@gatech.edu
lilia@cc.gatech.edu

GA Tech Mobile Robot Lab Study
gte246z@gatech.edu
lilia@cc.gatech.edu

GA Tech Mobile Robot Lab Study
gte246z@gatech.edu
lilia@cc.gatech.edu

GA Tech Mobile Robot Lab Study
gte246z@gatech.edu
lilia@cc.gatech.edu

GA Tech Mobile Robot Lab Study
gte246z@gatech.edu
lilia@cc.gatech.edu

GA Tech Mobile Robot Lab Study
gte246z@gatech.edu
lilia@cc.gatech.edu

GA Tech Mobile Robot Lab Study
gte246z@gatech.edu
lilia@cc.gatech.edu

GA Tech Mobile Robot Lab Study
gte246z@gatech.edu
lilia@cc.gatech.edu

APPENDIX X: DESCRIPTIVE STATISTICS FOR SEMANTIC DIFFERENTIAL SCALES, SEARCH-AND-RESCUE EXPERIMENT

X.1: UNDERSTANDABILITY

Table 59: Descriptive Statistics for *Understandability*

Dependent Variable	Condition	N	Mean	Std. Deviation	Std. Error
Clear	Control	14	4.0714	.91687	.24505
	Mood Only	14	3.6429	1.08182	.28913
	Mood and Emotion	15	3.5333	1.18723	.30654
	Total	43	3.7442	1.07111	.16334
Easy to Read	Control	14	3.9286	.91687	.24505
	Mood Only	14	3.7857	.97496	.26057
	Mood and Emotion	15	3.6667	.97590	.25198
	Total	43	3.7907	.94006	.14336
Consistent	Control	14	4.3571	.74495	.19910
	Mood Only	14	3.7857	1.05090	.28087
	Mood and Emotion	15	3.9333	1.22280	.31573
	Total	43	4.0233	1.03483	.15781
Easy to Understand	Control	14	3.7143	1.06904	.28571
	Mood Only	14	3.5000	1.22474	.32733
	Mood and Emotion	15	3.3333	.72375	.18687
	Total	43	3.5116	1.00882	.15384
Expressive	Control	14	3.0714	1.07161	.28640
	Mood Only	14	3.0000	.87706	.23440
	Mood and Emotion	15	3.6667	1.04654	.27021
	Total	43	3.2558	1.02569	.15642
Understandability (Overall)	Control	14	19.1429	3.00914	.80423
	Mood Only	14	17.7143	3.95024	1.05575
	Mood and Emotion	15	18.1333	2.92445	.75509
	Total	43	18.3256	3.29300	.50218

X.2: PERSUASIVENESS

Table 60: Descriptive Statistics for *Persuasiveness*

Dependent Variable	Condition	N	Mean	Std. Deviation	Std. Error
Compelling	Control	14	3.2857	1.06904	.28571
	Mood Only	14	4.0714	.61573	.16456
	Mood and Emotion	15	4.0000	.65465	.16903
	Total	43	3.7907	.86073	.13126
Appropriate	Control	14	4.0000	.67937	.18157
	Mood Only	14	4.3571	.74495	.19910
	Mood and Emotion	15	4.0000	1.13389	.29277
	Total	43	4.1163	.87856	.13398
Persuasive	Control	14	3.5714	1.28388	.34313
	Mood Only	14	3.9286	.61573	.16456
	Mood and Emotion	15	3.4667	1.30201	.33618
	Total	43	3.6512	1.11021	.16930
Sincere	Control	14	3.8571	1.02711	.27451
	Mood Only	14	4.5714	.51355	.13725
	Mood and Emotion	15	4.3333	.61721	.15936
	Total	43	4.2558	.78961	.12041
Convincing	Control	14	3.6429	1.21574	.32492
	Mood Only	14	4.1429	.66299	.17719
	Mood and Emotion	15	4.4000	.63246	.16330
	Total	43	4.0698	.91014	.13879
Persuasiveness (Overall)	Control	14	18.3571	4.12510	1.10248
	Mood Only	14	21.0714	1.73046	.46249
	Mood and Emotion	15	20.2000	3.52947	.91130
	Total	43	19.8837	3.41013	.52004

X.3: NATURALNESS

Table 61: Descriptive statistics for *Naturalness*

Dependent Variable	Condition	N	Mean	Std. Deviation	Std. Error
Natural	Control	14	2.9286	.82874	.22149
	Mood Only	14	3.2857	1.13873	.30434
	Mood and Emotion	15	3.3333	.61721	.15936
	Total	43	3.1860	.87982	.13417
Humanlike	Control	14	2.2143	.69929	.18689
	Mood Only	14	2.6786	.95287	.25467
	Mood and Emotion	15	2.5333	.99043	.25573
	Total	43	2.4767	.89279	.13615
Conscious	Control	14	3.0714	.73005	.19511
	Mood Only	14	3.7143	1.06904	.28571
	Mood and Emotion	15	4.0667	.88372	.22817
	Total	43	3.6279	.97647	.14891
Lifelike	Control	14	2.6429	1.00821	.26945
	Mood Only	14	2.7857	.89258	.23855
	Mood and Emotion	15	2.6667	1.11270	.28730
	Total	43	2.6977	.98886	.15080
Interactive	Control	14	3.3571	.84190	.22501
	Mood Only	14	3.5714	1.22250	.32673
	Mood and Emotion	15	3.8667	1.18723	.30654
	Total	43	3.6047	1.09413	.16685
Naturalness (Overall)	Control	14	14.2143	3.19082	.85278
	Mood Only	14	16.0357	4.06895	1.08747
	Mood and Emotion	15	16.4667	3.11372	.80396
	Total	43	15.5930	3.53260	.53872

X.4: SUBJECT NEGATIVE AFFECT

Table 62: Descriptive statistics for participants' *Negative Affect* and *Nervous* scores, before and after their interaction with the robot.

Dependent Variable	Condition	N	Mean	Std. Deviation	Std. Error
Subject Negative Affect, Before	Control	14	1.2208	.22471	.06006
	Mood Only	14	1.3098	.34072	.09106
	Mood and Emotion	15	1.3645	.26363	.06807
	Total	43	1.2999	.27989	.04268
Subject Negative Affect, After	Control	14	1.1753	.17994	.04809
	Mood Only	13	1.4266	.39169	.10864
	Mood and Emotion	15	1.5455	.62229	.16068
	Total	42	1.3853	.46085	.07111
Subject Nervous, Before	Control	13	1.6923	.63043	.17485
	Mood Only	13	1.3846	.65044	.18040
	Mood and Emotion	14	1.5000	.51887	.13868
	Total	40	1.5250	.59861	.09465
Subject Nervous, After	Control	14	1.5000	.51887	.13868
	Mood Only	13	1.8462	.68874	.19102
	Mood and Emotion	15	2.4667	1.18723	.30654
	Total	42	1.9524	.93580	.14440

APPENDIX Y: CONSENT FORM FOR EXTRAVERSION EXPERIMENT

GEORGIA INSTITUTE OF TECHNOLOGY
CONSENT TO BE A RESEARCH PARTICIPANT

Project Title: Human-Robot Interaction Study II

Investigators: *Arkin, R.C., Ph.D., Moshkina, L., and Park, S.*

Protocol and Consent Title: *Human-Robot Interaction Study II consent to be a research participant; Protocol # H10206; Date: July 2, 2010.*

You are being asked to be a volunteer in a research study.

Purpose:
The purpose of this study is to evaluate robotic behaviors in a human-robot interaction task. The results will help improve the design of robotic behaviors and promote robotics research. We expect to enroll 40-50 people in this study.

Exclusion/Inclusion Criteria:
Participants in this study must be over 18 years old and have at least high-school level education and English language proficiency.

Procedures:
You will be randomly assigned to one of two conditions. If you decide to be in the study, your participation will be limited to a single session which should not exceed 60 minutes. You will be invited to GA Tech Mobile Robot Laboratory, where you will first complete an initial personality questionnaire and perform two simple tasks. In one, you will be read a number of facts, and then take a short multiple-choice test on the facts. In the other, you will be asked to solve a simple math problem with pencil and paper. Once done, you will be asked to interact with a small humanoid robot and perform similar tasks during the interaction. Your interaction with the robot will last 3-5 minutes per task, and will take place in a mock-up demolition exhibit setting. With your permission, we will videotape your interaction with the robot. After each task, you will be asked to fill out a short questionnaire, and another one at the end of the study. These questionnaires will ask you about your impressions of the robot, and your demographics. You may stop at any time and for any reason.

Risks or Discomforts:
The risks involved are no greater than those involved in playing a video or a role-playing game.

Benefits:
You are not likely to benefit in any way from joining this study, however, you will have an opportunity to interact with a humanoid robot and expand your knowledge about robots in general.

Compensation to You:
For your time and effort, you will be compensated with a $15 Starbucks gift card. In case you do not finish the study, the gift card amount will be $5.

Confidentiality:
The following procedures will be followed to keep your personal information confidential in this study: The data collected about you will be kept private to the extent allowed by law. To protect your privacy, your records will be kept under a code number rather than by name. Your records, including videotapes, will be kept in locked files and only study staff will be allowed to look at them; the videotapes will transcribed and destroyed after the full data analysis is compete. Your name and any other fact that might point to you will not appear when results of this study are presented or published; we will ask your written permission for use of any videos or still photographs in demos and publications. Your privacy will be protected to the extent allowed by law. To make sure that this research is being carried out in the proper way, the Georgia Institute of Technology IRB may review study records. The Office of Human Research Protections may also look over study records during required reviews.

Costs to You:
There are no costs to you, other than your time, for being in this study.

In Case of Injury/Harm:
If you are injured as a result of being in this study, please contact Principal Investigator, Ronald C. Arkin, Ph.D., at telephone (404) 894- 8209. Neither the Principal Investigator nor Georgia Institute of Technology has made provision for payment of costs associated with any injury resulting from participation in this study.

Participant Rights:

- Your participation in this study is voluntary. You do not have to be in this study if you don't want to be.
- You have the right to change your mind and leave the study at any time without giving any reason and without penalty.
- Any new information that may make you change your mind about being in this study will be given to you.
- You will be given a copy of this consent form to keep.
- You do not waive any of your legal rights by signing this consent form.

Conflict of Interest:
None.

Questions about the Study:
If you have any questions about the study, you may contact Dr. Ronald C. Arkin at telephone (404) 894- 8209 or at arkin@cc.gatech.edu.

Questions about Your Rights as a Research Participant:

If you have any questions about your rights as a research participant, you may contact

Ms. Melanie Clark, Georgia Institute of Technology
Office of Research Compliance, at (404) 894-6942.

APPENDIX Y

If you sign below, it means that you have read the information given in this consent form, and you would like to be a volunteer in this study.

____
Participant Name (printed)

____ ______________
Participant Signature Date

____ ______________
Signature of Person Obtaining Consent Date

Consent Form approved by Georgia Tech IRB from September 8, 2010 to July 20, 2011

APPENDIX Z: MEMORY RETENTION TASK

Z.1: BASELINE MEMORY TASK

VIDEO PRESENTATION

Housing Construction – Baseline Task

In the following presentation, I will tell you how residential houses are built in the United States. Please listen carefully, as there will be a short quiz at the end. All right. You may know that in the US there are more than 100 million housing units, and the majority of them are "single family dwellings," or houses. One of the amazing things about American homes is that the huge majority of them are built using completely standardized building practices. One reason for this consistency is a set of uniform building codes that apply across the country. Another reason is cost -- the techniques used to build homes produce reliable housing quickly at a low cost. If you ever watch any house being built, you will find that it roughly goes through 7 general steps, which include a number of sub-steps. I will take you through these 7 steps briefly.

The first step includes site preparation and foundation construction. The site-preparation crew typically arrives on the site with a backhoe and/or bulldozer. The crew's job is to clear the site of any trees, rocks and debris, level the site if necessary and dig for the foundation being built. Slabs, basements and crawl spaces are the three main foundation systems used on houses. The slab is a flat concrete pad poured directly on the ground. It takes very little site preparation, very little formwork for the concrete and very little labor to create. A house with a **basement** starts with a hole about 8 feet deep. At the bottom of the hole is a concrete slab, and then concrete or cinder-blocks form the outer walls of the basement. Finally, a **crawl space** has several advantages over basements and slabs:

- It gets the house up off the ground (especially important in damp, cold or termite-prone areas).
- It is a lot less expensive than a basement and comparable in price to a slab.

The second step is framing. Framing starts by building the floor; it can either sit directly on the foundation, or on “joists”, which are brick posts used to lift the floor higher up. Once the floor framing is complete, it is covered with 1/2-inch plywood or OSB (which stands for oriented strand board). Next, the framing crew starts on the walls. Walls are assembled on the floor, and then raised into place. Walls are usually made of lumber and covered on the outside with an OSB sheathing. Using plywood or OSB as the sheathing gives the wall rigidity and strength. All of the exterior walls go up following this same basic pattern. In the corners, the **top plate** on one wall overlaps the top plate of the next, and the walls are nailed together to bind the corner. Then the interior walls go up, fitting into the top plates of the exterior walls. So, framing is the second step of the construction process.

Next comes roofing, doors and windows. **In modern construction, trusses** are used for the roof framing. Trusses are pre-fabricated, triangulated wooden structures used to support the roof. Trusses are quite common these days because they have five big advantages from the builder's standpoint: they are strong, less expensive, can be custom-built, they go up quickly, and spread the weight evenly. The main disadvantage is that they don't allow for attic space. The trusses are fist stacked on top of the walls, either by hand or with a crane; then they are tied to the walls with small metal plates. Once the trusses are up, the roof is covered in plywood or OSB, which gives the roof tremendous rigidity. The windows and doors are usually prefabricated, arrive in one shipment and are unloaded from the truck into a stack. Plastic stripping is stapled to the inside of all window and door openings, and the windows are placed in each rough opening and stapled in place on the outside. Once all the doors and windows are in, the roof is covered with shingles and equipped with ridge vents for better circulation.

The fourth step of house construction is siding – this is the last exterior step. The siding can come in many different variations, and vinyl is one of the most popular ones. It is made from thin, flexible sheets of plastic about 2 millimeters thick, pre-colored and bent into shape during manufacturing. The sheets are 12 feet long and about a foot high. You start at the bottom and the sheets **interlock** into each other as you go up. Once the siding is in, the house is "dried in" – meaning that it is completely protected from rain.

The next 3 steps are all interior. First, rough plumbing and rough electrical need to go in. Both of these are rather complicated, and have a lot of subtleties and nuances. Typically, rough plumbing involves installing all sewer lines and vents as well as all water supply lines for each fixture: sinks, bathtubs, toilets and washers. In the crawl space, the supply lines all branch off from common pipes running the length of the house. The sewer lines all join together and then exit out the back of the house, ready for connection to the septic tank. For rough electrical, the electrician first puts all of the boxes for electrical outlets, lights and switches; then he runs wires from the fuse box to each box and between boxes. A lot of drilling is necessary, both down into the crawl space and up into the ceiling.

The next step is the insulation and dry wall. The purpose of insulation is to lower the heating and cooling costs for the house by limiting heat transfer through the walls and the ceiling. The insulation process starts by installing **foam channels** in the eaves; these channels guarantee that air will be able to flow from the interior vents to the ridge vents. Once the insulation is in, drywall goes up. Drywall (also known as "plaster board" and by the trade name "Sheetrock") is a half-inch layer of plaster or gypsum sandwiched between two thick sheets of paper. It is remarkably solid, and also remarkably heavy. The drywallers put up all of the drywall in a day and **tape** it the next day.

Finally, all that remains are the finishing steps. These include: installing heating and air conditioning, finishing electrical and plumbing, installing kitchen and bathroom cabinets and counters, wall trim and painting, and, finally, carpeting and tiling the floors. Voila – now you have a fully finished house!

Thank you very much for listening to this presentation. Now, please take the short quiz in front of you. Let me know when you are done. Once you are done, please fill out a brief questionnaire, and call me. I will then take you to the experiment area.

BASELINE QUIZ

Please recall the facts the experimenter recited for you. In your answers, please refer to the content specifically presented by the experimenter, not the general knowledge you may have of the house construction process.

1) The main reason(s) for standardized building practices is/are:
    a) Cost
    b) Laws
    c) Cross-country applicability
    d) Safety
    e) a) and c)

2) OSB stands for:
    a) Obfuscated striated board
    b) Oriented strand board
    c) Orange strange ball
    d) Oblique strand board

3) The purpose of using plywood in framing and roofing is :
    a) Aesthetics
    b) Safety
    c) Cost
    d) Strength

4) In roofing, the main disadvantage of trusses is:
    a) They don’t support the weight evenly
    b) They are expensive
    c) They don’t allow for attic space
    d) They are too long in length
    e) They are not strong enough for larger buildings

5) The finishing step includes:
    a) Cabinetry;
    b) Painting;
    c) Mounting dry wall;
    d) Finishing up electrical;
    e) a), b) and d)
    f) a), c) and d)
    g) All of the above

Z.2: EXPERIMENTAL MEMORY RETENTION TASK

PRESENTATION, EXTRAVERTED

Welcome back, I've been waiting for you! You know, at the end of this presentation you will be asked to take a short quiz. So please pay close attention while I tell you all I know about the building demolition process. You know, sometimes old buildings are not safe anymore, and need to be replaced by new, safer and more reliable structures. When such buildings are relatively small, say, 3 to 5 stories high, sledgehammers, excavators and wrecking balls are perfectly capable of reducing them to rubble. However, when it's a huge 20-story skyscraper we are talking about, such low-tech means are not very effective. For demolition of large buildings, explosives are used to collapse them, and that's a lot more exciting than sledgehammers! The basic idea of explosive demolition is quite simple: If you remove the support structure of a building at a certain point, the section of the building above that point will fall down on the part of the building below that point. If this upper section is heavy enough, it will collide with the lower part with sufficient force to cause significant damage. Boom! However, the explosives are just the trigger for the demolition. It's gravity that brings the building down.

a. Although the idea may be simple, the process itself is not, and the preparations may take weeks, or even months. From the beginning to the end, explosive demolition of a building takes 5 steps, and I will take you through these exciting steps. The first step is planning the demolition. First, the blaster team examines architectural blueprints to determine how the building was put together. After that, blaster crew tours the building several times to jot notes about the support structure, and then puts together a plan of attack, based on prior experience. This includes what explosives to use, where to position them in the building, and how to time their detonations. Sometimes, they develop approximate 3-D models to test the plan in a virtual world. So, planning is the first step of the explosive demolition process.
b. Once the planning is done, the second step is preparing the building. During building preparation, the crew first clears out the debris, and then takes out non-load bearing walls within the building. After that, the load-bearing columns are loaded with explosives. And here it is starting to get exciting! Dynamite is used for concrete columns, and specialized explosive material for steel columns. Once the explosives are loaded, the team arranges blasting caps on the columns. These are small amounts of explosive material, connected to some sort of fuse. And that fuse better not be too short, right? To precisely control the timing of the individual explosions, the blasting caps are configured with simple delay mechanisms. Finally, to reduce the amount of flying debris, the crew wraps chain-link fence and geotextile fabric around each column, where explosives are positioned. So, building preparation is the second step.
c. Once the building is carefully prepared for explosive demolition, the next, very important, step is insuring safety. First local authorities and neighboring businesses need to be assured that the demolition won't seriously damage nearby structures. Explosions can certainly be scary! Next, the last check of the explosives needs to be performed. Nothing should be left to chance. And, of course, the main thing is to make sure

the building and the area surrounding it are completely clear. For this, anyone in the dangerous area needs to be evacuated. Surprisingly, implosion enthusiasts, sometimes try to sneak past barriers for a closer view of the blast, despite the obvious risks. Can you believe that? So, insuring safety is the third step.

d. Now that everyone is cleared out of the area, the next step is the execution, or blasting itself. For this step, the crew retreats to the detonator controls, and begins the countdown. The blasters may sound a siren at the 10-minute, five-minute and one-minute mark, to let everyone know when the building will be coming down. Then the button is pressed, and the explosions start. Boom! Boom! Typically, the actual demolition only takes a few seconds – but what a sight! To many onlookers, the speed of destruction is the most incredible aspect of an explosive demolition. How can a building that took months or years to build, collapse into a pile of rubble as if it were a sand castle? So, the actual explosive demolition is the fourth step of the entire process.
e. Finally, when all the excitement is over, the last step is surveying the results. Following the blast, a cloud of dust billows out around the wreckage, enveloping nearby spectators. After the cloud has cleared, the blasters survey the scene. At this stage, it is crucial to confirm that all of the explosives were detonated and to remove any explosives that did not go off. Accidents are certainly not welcome. Most of the time, experienced blasters bring buildings down exactly as planned, and everything goes off without a hitch. Damage to nearby structures, even ones immediately adjacent to the blast site, is usually limited to a few broken windows. So, surveying the results is the final step of the explosive demolition process.

Now you know what happens when large, multi-story buildings need to be demolished. Isn't that exciting? Thank you very much for listening. Now we will see how well you remember what I talked about! Please go over there, to my assistant, Lilia, and take a quick test. Bye-bye for now.

PRESENTATION, INTROVERTED

Welcome back. At the end of this presentation you will take a short quiz. Please pay attention while I tell you about the building demolition process. Sometimes old buildings are not safe anymore, and need to be replaced by new, safer and more reliable structures. When such buildings are relatively small, 3 to 5 stories high, sledgehammers, excavators and wrecking balls are capable of reducing them to rubble. However, when it's a 20-story skyscraper, such low-tech means are not very effective. For demolition of large buildings, explosives are used to collapse them. The basic idea of explosive demolition is simple: If you remove the support structure of a building at a certain point, the section of the building above that point will fall down on the part of the building below that point. If this upper section is heavy enough, it will collide with the lower part with sufficient force to cause significant damage. The explosives are just the trigger for the demolition. It's gravity that brings the building down.

a. Although the idea may be simple, the process itself is not, and the preparations may take months. From the beginning to the end, explosive demolition of a building takes 5 steps, and I will take you through these steps. The first step is planning the demolition, during which the blaster team maps each element of the process ahead of time. First, they examine architectural blueprints to determine how the building was put together. Blaster crew tours the building several times to jot notes about the support structure, and puts together plan of attack, based on prior experience. This includes what explosives to use, where to position them in the building, and how to time their detonations. Sometimes, they develop 3-D models to test the plan in a virtual world. Planning is the first step of the explosive demolition process.

b. Once the planning is done, the second step is preparing the building. During building preparation, the crew first clears out the debris, and takes out non-load bearing walls within the building. Then, the load-bearing columns are loaded with explosives. Dynamite is used for concrete columns, and specialized explosive material for steel columns. Once the explosives are loaded, the team arranges blasting caps on the columns. These are small amounts of explosive material, connected to some sort of fuse. To precisely control the timing of the individual explosions, the blasting caps are configured with simple delay mechanisms. Finally, to reduce the amount of flying debris, the crew wraps chain-link fence and geotextile fabric around each column, where explosives are positioned. Building preparation is the second step.
c. Once the building is prepared for explosive demolition, the next step is insuring safety. First, local authorities and neighboring businesses need to be assured that the demolition won't seriously damage nearby structures. Next, the last check of the explosives needs to be performed. Finally, the team makes sure the building and the area surrounding it are completely clear. For this, anyone in the dangerous area needs to be evacuated. Sometimes, implosion enthusiasts try to go past barriers for a closer view of the blast, despite the obvious risks. Insuring safety is the third step.
d. Now that everyone is cleared out of the area, the next step is the execution, or blasting itself. For this step, the crew retreats to the detonator controls, and begins the countdown. The blasters may sound a siren at the 10-minute, five-minute and one-minute mark, to let everyone know when the building will be coming down. Then the button is pressed, and the explosions start. Typically, the actual demolition only takes a few seconds. To many onlookers, the speed of destruction is the most incredible aspect of an explosive demolition. How can a building that took months and months to build, collapse into a pile of rubble? The actual explosive demolition is the fourth step of the process.
e. Finally, the last step is surveying the results. Following the blast, a cloud of dust billows out around the wreckage, covering nearby spectators. After the cloud has cleared, the blasters survey the scene. At this stage, it is crucial to confirm that all of the explosives were detonated and to remove any explosives that did not go off. Most of the time, experienced blasters bring buildings down exactly as planned. Damage to nearby structures, even ones immediately adjacent to the blast site, is usually limited to a few broken windows. Surveying the results is the final step of the process.

Now you know what happens when large, multi-story buildings need to be demolished. Thank you for listening. Now we will see how well you remember what I talked about! Please go to my assistant Lilia and take a quick test. Bye-bye.

EXPERIMENTAL QUIZ

Please recall the facts the robot guide recited for you. In your answers, please refer to the content specifically presented by the robot, not the general knowledge you may have of the demolition process.

1) The main force in explosive demolition is/are:
   a) The explosives
   b) The gravity
   c) The blaster crews

2) The five steps of explosive demolition include all of these except :
   a) Planning the demolition
   b) Blasting itself
   c) Purchasing explosive materials
   d) Surveying the results
   e) Preparing the building

3) The planning step does not include:
   a) Examining architectural blueprints to determine how the building was put together
   b) Touring the building several times to jot notes about the support structure
   c) Putting together a plan of attack, based on prior experience: what explosives to use, where to position them in the building, and how to time their detonations
   d) Testing the explosives on a smaller structure, to make sure everything is as it should be

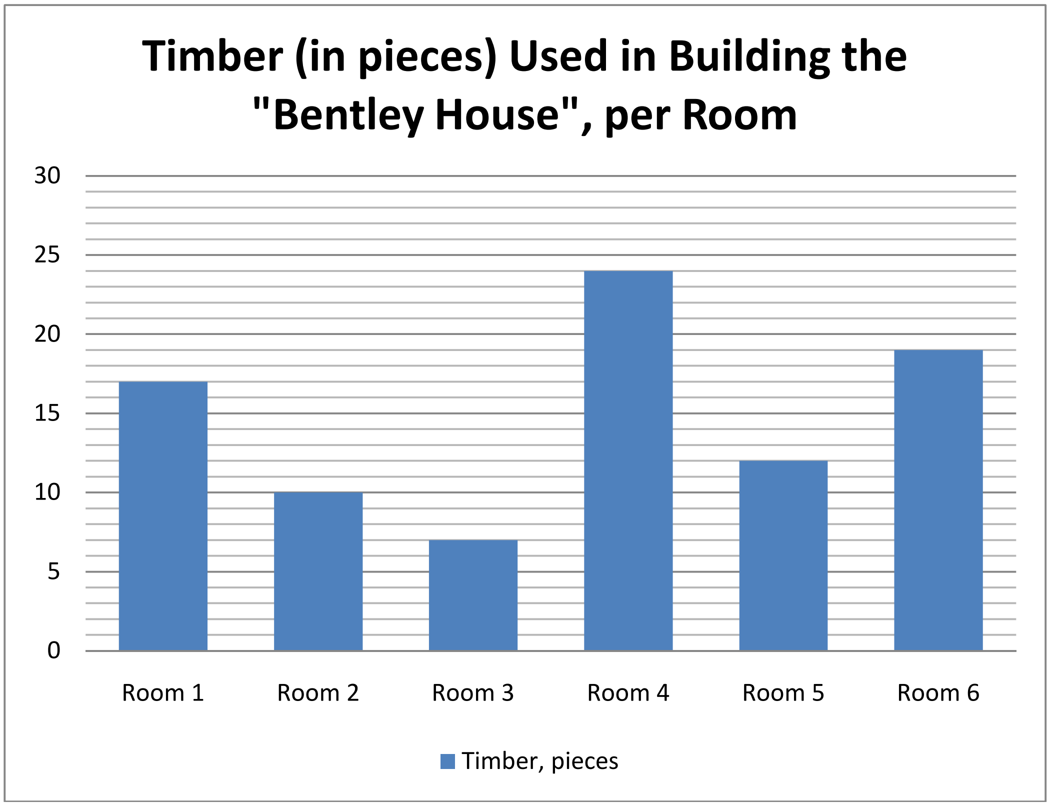
4) Dynamite is used to:
   a) Demolish steel columns
   b) Demolish non-load bearing walls
   c) Demolish concrete columns
   d) a) and c)
   e) All of the above

5) The typical damage during building demolition usually:
   a) Involves surrounding buildings significantly;
   b) May result in human casualties;
   c) Is limited to a few broken windows in the surrounding area;
   d) Is limited to a few cracks in the surrounding buildings;
   e) c) and d)

APPENDIX AA: MATH PROBLEM

AA.1: BASELINE MATH PROBLEM

Please solve the following multiple-choice math problem. Feel free to use this paper as scratch paper for your calculations.

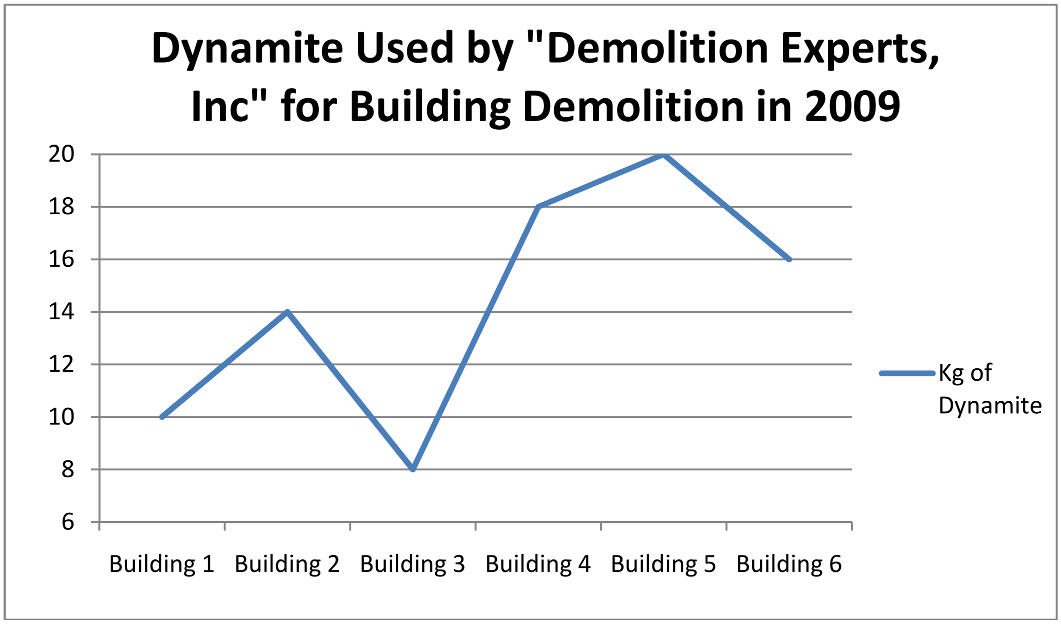


The figure above shows the number of pieces of timber that Expert Housing, Inc used per room in building their latest project, the “Bentley House”. What is the average (arithmetic mean) of the timber pieces the company used per room for the entire “Bentley House”?

a) 11.61
b) 12.73
c) 13.00
d) 14.33
e) 14.83

AA.2: EXPERIMENTAL MATH PROBLEM

Please solve the following multiple-choice math problem. Feel free to use this paper as scratch paper for your calculations.



The figure above shows the amount of dynamite, in Kg, that Demolition Experts, Inc used per building in their explosive demolition projects in 2009. What is the average (arithmetic mean) of the amount of dynamite the company used per building for the entire year of 2009?

a) 11.66
b) 12.67
c) 13.00
d) 14.33
e) 15.33

APPENDIX AA

ROBOT'S INSTRUCTIONS: EXTRAVERTED

Hello again, I have been waiting for you! Now, you will have a first-hand experience with the types of task, members of demolition crews might face. Please follow my instructions carefully. For this task, you have been given a math problem on a piece of paper. Please read it over, and solve it using the paper and pencil you were given. Please try to be as precise as possible. I am sure you will enjoy it. Tell my assistant, Lilia, and not me, when you are all done. Good luck! You can start now.

So, how is it going?

You know, this exhibit is kind of small?

Small talk

By the way, have you noticed that high voltage sign?

I bet the weather is pretty good today!

I am afraid the time allotted to this task is up. Please take your answers to my assistant, Lilia, and thank you for your participation!

ROBOT'S INSTRUCTIONS: INTROVERTED

Hello again. Now, you will have a first-hand experience with the types of task, members of demolition crews might face. Please follow my instructions carefully. For this task, you have been given a math problem. Please read it over, and solve it using the paper and pencil you were given. Please try to be as precise as possible. Tell my assistant Lilia, and not me, when you are done. You can start now.

How is it going? Small talk

The time allotted to this task is up. Please take your answers to my assistant, Lilia. Thank you.

APPENDIX BB: ROBOT'S GREETING FOR EXTRAVERSION EXPERIMENT

EXTRAVERTED GREETING:

Hello, Visitor! Thank you very much for coming today. I hope you have a lot of fun during this little tour. My name is Nao, and I am a humanoid robot manufactured by a French company named Aldebaran Robotics. I was brought here all the way from France, about 3 months ago. Today, I will be your guide on this exhibit. This exhibit was designed to show visitors, like you, what goes on during building demolition. While you are here, I will tell you how, blaster, crews, demolish buildings with explosives, and you will also have a chance to have first-hand experience in what kind of planning goes into demolition process. All right, now please take a couple of minutes to look around the exhibit. When you are done, let the experimenter, Lilia, know, and then come back to see me! Then we will be ready to begin!

INTROVERTED GREETING:

Hello, Visitor! Thank you for coming today. My name is Nao, and I am a humanoid robot manufactured by a French company named Aldebaran Robotics. I was brought here from France, 3 months ago. Today, I will be your guide on this exhibit. This exhibit was designed to show visitors what goes on during building demolition. I will tell you how, blaster, crews, demolish buildings with explosives, and you will also have first-hand experience in what kind of planning goes into demolition process. Now please take a couple of minutes to look around. When you are done, let the experimenter, Lilia, know, and then come back here. We will be ready to begin.

APPENDIX CC: POST-QUIZ QUESTIONNAIRE

Please reflect back on your interaction with the robot during the task you have just completed as you consider the questions below.

For the next 3 questions, please rate your impressions of the robot DURING THIS TASK by circling the most appropriate number on the scale:

1. In your opinion, FOR THIS TASK, the robot's behavior was:

Inappropriate				Appropriate
1	2	3	4	5

Wrong for Task				Right for Task
1	2	3	4	5

Ill-suited				Well-suited
1	2	3	4	5

Improper				Proper
1	2	3	4	5

Mismatched				Matched to Task
1	2	3	4	5

2. In your opinion, YOUR PRESENCE during the interaction with the robot was:

Unwelcome				Welcome
1	2	3	4	5

Undesired				Desirable
1	2	3	4	5

Disliked				Liked
1	2	3	4	5

Tolerated				Encouraged
1	2	3	4	5

Unwanted				Wanted
1	2	3	4	5

Participant #

TURN OVER →

3. In your opinion, THE FACTS the robot presented to you were:

Boring				Interesting
1	2	3	4	5

Not Fun				A Lot of Fun
1	2	3	4	5

Useless				Useful
1	2	3	4	5

Dull				Exciting
1	2	3	4	5

Tedious				Entertaining
1	2	3	4	5

APPENDIX DD: POST-MATH QUESIONNAIRE

Please reflect back on your interaction with the robot during the task you have just completed as you consider the questions below.

For the next 3 questions, please rate your impressions of the robot DURING THIS TASK by circling the most appropriate number on the scale:

1. In your opinion, FOR THIS TASK, the robot's behavior was:

Inappropriate				Appropriate
1	2	3	4	5
Wrong for Task				Right for Task
1	2	3	4	5
Ill-suited				Well-suited
1	2	3	4	5
Improper				Proper
1	2	3	4	5
Mismatched				Matched to Task
1	2	3	4	5

2. In your opinion, during this task, the ROBOT WAS:

Distracting				Easy to Tune Out
1	2	3	4	5
Interfering				Minding its own business
1	2	3	4	5
Annoying				Inoffensive
1	2	3	4	5
Irritating				Undemanding
1	2	3	4	5
Bothersome				Quiet
1	2	3	4	5

Participant #

TURN OVER →

3. The MATH PROBLEM you have just solved was:

Hard — Easy

1 2 3 4 5

Complicated — Simple

1 2 3 4 5

Demanding — Undemanding

1 2 3 4 5

Long — Short

1 2 3 4 5

Complex — Basic

1 2 3 4 5

APPENDIX EE: NASA TLX SCALE

PLEASE RATE THE ***MATH*** TASK YOU HAVE JUST COMPLETED ON THE FOLLOWING SCALES. CIRCLE OR CROSS THE MOST APPROPRIATE MARK FOR EACH SCALE. USE THE DEFINITIONS ON THE BACK IF NEEDED.

Mental Demand — How mentally demanding was the task?

Very Low — Very High

Physical Demand — How physically demanding was the task?

Very Low — Very High

Temporal Demand — How hurried or rushed was the pace of the task?

Very Low — Very High

Performance — How successful were you in accomplishing what you were asked to do?

Perfect — Failure

Effort — How hard did you have to work to accomplish your level of performance?

Very Low — Very High

Frustration — How insecure, discouraged, irritated, stressed, and annoyed wereyou?

Very Low — Very High

PARTICIPANT #:

RATING SCALE DEFINITIONS

Title	Endpoints	Descriptions
MENTAL DEMAND	Low/High	How much mental and perceptual activity was required (e.g., thinking, deciding, calculating, remembering, looking, searching, etc.)? Was the task easy or demanding, simple or complex, exacting or forgiving?
PHYSICAL DEMAND	Low/High	How much physical activity was required (e.g., pushing, pulling, turning, controlling, activating, etc.)? Was the task easy or demanding, slow or brisk, slack or strenuous, restful or laborious?
TEMPORAL DEMAND	Low/High	How much time pressure did you feel due to the rate or pace at which the tasks or the task elements occurred? Was the pace slow and leisurely or rapid and frantic?
PERFORMANCE	Good/Poor	How successful do you think you were in accomplishing the goals of the task set by the experimenter (or yourself)? How satisfied were you with your performance in accomplishing these tasks?
EFFORT	Low/High	How hard did you have to work (mentally and physically) to accomplish your level of performance?
FRUSTRATION LEVEL	Low/High	How insecure, discouraged, irritated, stressed and annoyed vs. secure, gratified, content, relaxed, and complacent did you feel during the task?

SOURCE-OF-WORKLOAD COMPARISON CARDS

These squares were cut into cards, each containing two categories from the TLX scale, and the participants were asked to circle the category they found more relevant/demanding during the performance of the task, for each of the cards. The score was calculated by counting the number of times each category was circled, and then used as a weight for that category to determine the overall workload score.

Effort or Performance	Temporal Demand or Frustration	Temporal Demand or Effort	Physical Demand or Frustration
Performance or Frustration	Physical Demand or Temporal Demand	Physical Demand or Performance	Temporal Demand or Mental Demand
Frustration or Effort	Performance or Mental Demand	Performance or Temporal Demand	Mental Demand or Effort
Mental Demand or Physical Demand	Effort or Physical Demand	Frustration or Mental Demand	

APPENDIX FF: ROBOT PERSONALITY QUESTIONNAIRE, TRAIT EXPERIMENT

How Accurately Can You Describe The Robot?

Please use this list of common traits to describe the robot as accurately as possible. Describe the robot based on your observations during the tasks you performed. Before each trait, please write a number indicating how accurately that trait describes the robot, using the following rating scale:

Extremely Inaccurate ……………… 1

Very Inaccurate ……………………. 2

Moderately Inaccurate …………….. 3

Slightly Inaccurate ………………….4

Neither Inaccurate or Accurate ……. 5

Slightly Accurate ………………….. 6

Moderately Accurate ………………. 7

Very Accurate …………………….... 8

Extremely Accurate ………………...9

____ Bashful	____ Energetic	____ Moody	____ Systematic
____ Bold	____ Envious	____ Organized	____ Talkative
____ Careless	____ Extraverted	____ Philosophical	____ Temperamental
____ Cold	____ Fretful	____ Practical	____ Touchy
____ Complex	____ Harsh	____ Quiet	____ Uncreative
____ Cooperative	____ Imaginative	____ Relaxed	____ Unenvious
____ Creative	____ Inefficient	____ Rude	____ Unintellectual
____ Deep	____ Intellectual	____ Shy	____ Unsympathetic
____ Disorganized	____ Jealous	____ Sloppy	____ Warm
____ Efficient	____ Kind	____ Sympathetic	____ Withdrawn

Bashful: socially shy or timid; diffident; self-conscious.
Bold: fearless before danger; intrepid; impudent; presumptuous.
Careless: Marked by lack of attention, consideration, forethought or thoroughness; not careful.
Cold: Marked by a lack of the warmth of normal human emotion, friendliness, or compassion.
Complex: Hard to separate, analyze, or solve; complicated.
Cooperative: marked by a willingness and ability to work with others.
Creative: Characterized by originality and expressiveness; imaginative.
Deep: Of penetrating intellect; wise.
Disorganized: To be into utter disorder; disarrange.
Efficient: Exhibiting a high ratio of output to input; effective.
Energetic: Operating with or marked by vigor or effect; vigorous.
Envious: Painfully desirous of another's advantages; jealous; covetous.
Extraverted: a gregarious and unreserved person; outgoing.
Fretful: Marked by worry and distress; inclined to be vexed or troubled.
Harsh: Unpleasantly stern; severe.
Imaginative: Created by, indicative of, or characterized by imagination; having no truth; false.
Inefficient: Wasteful of time, energy, or materials; lacking the ability or skill to perform effectively; incompetent:
Intellectual: Having or showing intellect, especially to a high degree; intelligent
Jealous: Fearful or wary of being supplanted; apprehensive of losing affection or position; resentful or bitter in rivalry; envious.
Kind: Of a friendly, generous, or warm-hearted nature; considerate.
Moody: Subject to depression or moods; expressive of a mood; temperamental.
Organized: Methodical and efficient in arrangement or function; orderly.
Philosophical: Characteristic of a philosopher, as in equanimity, enlightenment, and wisdom.
Practical: Concerned with actual use or practice; useful.
Quiet: Restrained in style; understated; making little or no noise.
Relaxed: To be less restrained or tense; easy and informal in manner.
Rude: Ill-mannered; discourteous; uncouth.
Shy: Marked by reserve or diffidence; reserved; wary.
Sloppy: Marked by a lack of neatness or order; untidy.
Sympathetic: Expressing or feeling compassion or friendly fellow feelings.
Systematic: Characterized by order and planning; orderly.
Talkative: Full of trivial conversation; loquacious; garrulous; voluble.
Temperamental: Marked by excessive sensitivity and impulsive changes of mood.
Touchy: marked by readiness to take offense on slight provocation; sensitive.
Uncreative: opposite of creative (see above)
Unenvious: opposite of envious (see above)
Unintellectual: opposite of intellectual (see above)
Unsympathetic: opposite of sympathetic (see above)
Warm: Marked by or revealing friendliness or sincerity; loving; kind.
Withdrawn: Not friendly or sociable; aloof; detached; emotionally unresponsive.

APPENDIX GG: POST-QUESTIONNAIRE FOR TRAIT EXPERIMENT

Please rate your impressions of the robot during the entire interaction by circling the most appropriate number on the scale:

1. In your opinion, the robot APPEARED:

Fake				Natural
1	2	3	4	5

Machinelike				Humanlike
1	2	3	4	5

Unconscious				Conscious
1	2	3	4	5

Artificial				Lifelike
1	2	3	4	5

Inanimate				Animate
1	2	3	4	5

2. In your own words, please briefly describe your interaction with the robot. Please compare the robot's behavior and your perception of it during the two tasks.

Participant #

TURN OVER →

3) Finally, please describe your expectations of robots before today's experiment, and how they compare to your impressions of the robot today.

APPENDIX HH: DEMOGRAPHICS QUESTIONNAIRE FOR TRAIT EXPERIMENT

Demographics Questionnaire

1. What is your gender?
    a.) Female
    b.) Male

2. What is your age?
    a.) Under 20 years old
    b.) Between 20 and 30
    c.) Between 30 and 40
    d.) Between 40 and 50
    e.) 50 or older

3. What is the highest level of education you've achieved? What was your **major**?
    a.) High School
    b.) Currently working on my Bachelor's ____________________
    c.) Bachelor's ____________________
    d.) Master's ____________________
    e.) Currently working on my Ph.D. ____________________
    f.) Ph.D. ____________________
    g.) Other ____________________

4. Is English your native language?
    a.) Yes
    b.) No

5. Do you describe yourself as technical (having extensive experience or interest in a technical field, such as engineering, computing, math, etc)?
    a.) Yes
    b.) Somewhat
    c.) No

6. What is your level of computer experience?
    a.) None: Never used a computer before
    b.) Limited: Occasionally use a computer for tasks like e-mail, web browsing or word processing
    c.) User Level: Regularly use a computer for tasks like e-mail, web browsing or word processing
    d.) Advanced User: Have downloaded and installed at least one program from the Internet
    e.) Programmer Level: Some programming experience
    f.) Advanced Programmer: Extensive training or experience in multiple programming languages

7. Have you ever interacted with robots? Please circle all that apply.
    a.) Never
    b.) Very limited interaction
    c.) Interaction experience with military robots
    d.) Interaction experience with industrial robots
    e.) Interaction experience with mobile robots
    f.) Interaction experience with entertainment/educational robots
    g.) Interaction experience with humanoid robots
    i.) Interaction experience with humanoid robot Nao
    j.) Other – please specify ____

8. Have you participated in another experiment with the same robot in this lab?
    a.) Yes
    b.) No

APPENDIX II: RELIABILITY AND DESCRIPTIVE STATISTICS FOR SEMANTIC DIFFERENTIAL SCALES (SCALE AND SUBSCALE LEVEL)

MOOD EXPERIMENT SCALES

Table 63: Reliability and Descriptive Statistics for Understandability

Scale: Understandability

Mean	Variance	Std. Deviation	N of Items
18.3256	10.844	3.29300	5

Item Statistics

	Mean	Std. Deviation	N of cases
Clear	3.7442	1.07111	43
Easy to Read	3.7907	.94006	43
Consistent	4.0233	1.03483	43
Easy to Understand	3.5116	1.00882	43
Expressive	3.2558	1.02569	43
Cronbach's Alpha: 0.654			

Table 64: Reliability and Descriptive Statistics for Persuasiveness

Scale: Persuasiveness

Mean	Variance	Std. Deviation	N of Items
19.8837	11.629	3.41013	5

Item Statistics

	Mean	Std. Deviation	N of cases
Compelling	3.7907	.86073	43
Appropriate	4.1163	.87856	43
Persuasive	3.6512	1.11021	43
Sincere	4.2558	.78961	43
Convincing	4.0698	.91014	43
Cronbach's Alpha: 0.799			

Table 65: Reliability and Descriptive Statistics for Naturalness

Scale: Naturalness (Mood and Emotion)

Mean	Variance	Std. Deviation	N of Items
15.5930	12.479	3.53260	5

Item Statistics

	Mean	Std. Deviation	N of cases
Natural	3.1860	.87982	43
Humanlike	2.4767	.89279	43
Conscious	3.6279	.97647	43
Lifelike	2.6977	.98886	43
Interactive	3.6047	1.09413	43
Cronbach's Alpha: 0.779			

EXTRAVERSION EXPERIMENT SCALES

Table 66: Reliability and Descriptive Statistics for Quiz Appropriateness

Scale: Quiz Appropriateness

Mean	Variance	Std. Deviation	N of Items
19.4667	18.947	4.35283	5

Item Statistics

	Mean	Std. Deviation	N of cases
Appropriate (quiz)	4.0333	.96431	30
Right for Task (quiz)	3.7667	1.00630	30
Well-suited (quiz)	3.8667	.97320	30
Proper (quiz)	4.0333	.96431	30
Matched to Task (quiz)	3.7667	1.10433	30
Cronbach's Alpha: 0.918			

Table 67: Reliability and Descriptive Statistics for Math Appropriateness

Scale: Math Appropriateness

Mean	Variance	Std. Deviation	N of Items
16.8571	36.497	6.04130	5

Item Statistics

	Mean	Std. Deviation	N of cases
Appropriate (math)	3.5357	1.31887	28
Right for Task (math)	3.3929	1.28638	28
Well-suited (math)	3.3571	1.25357	28
Proper (math)	3.4643	1.34666	28
Matched to Task (math)	3.1071	1.22744	28
Cronbach's Alpha: 0.966			

Table 68: Reliability and Descriptive Statistics for Welcome

Scale: Welcome

Mean	Variance	Std. Deviation	N of Items
19.2667	14.961	3.86793	5

Item Statistics

	Mean	Std. Deviation	N of cases
Welcome	4.0333	.88992	30
Desirable	3.7667	.89763	30
Liked	3.7667	.85836	30
Encouraged	3.9667	.92786	30
Wanted	3.7333	.90719	30
Cronbach's Alpha: 0.914			

Table 69: Reliability and Descriptive Statistics for Appeal

Scale: Appeal

Mean	Variance	Std. Deviation	N of Items
16.4333	16.806	4.09948	5

Item Statistics

	Mean	Std. Deviation	N
Interesting	3.8667	1.19578	30
Fun	3.0333	.96431	30
Useful	3.5667	1.13512	30
Exciting	3.0000	.90972	30
Entertaining	2.9667	.96431	30
Cronbach's Alpha: 0.848			

Table 70: Reliability and Descriptive Statistics for Unobtrusiveness

Scale: Unobtrusiveness

Mean	Variance	Std. Deviation	N of Items
17.3571	29.201	5.40380	5

Item Statistics

	Mean	Std. Deviation	N of cases
Easy to Tune Out	3.5357	1.45251	28
Minding its Own Business	3.2500	1.26564	28
Inoffensive	3.5000	1.17063	28
Undemanding	3.6786	1.12393	28
Quiet	3.3929	1.10014	28
Cronbach's Alpha: 0.927			

Table 71: Reliability and Descriptive Statistics for Ease

Scale: Ease

Mean	Variance	Std. Deviation	N of Items
23.1786	6.819	2.61128	5

Item Statistics

	Mean	Std. Deviation	N of cases
Easy	4.6429	.62148	28
Simple	4.7500	.51819	28
Undemanding	4.6786	.61183	28
Short	4.4286	.83571	28
Basic	4.6786	.61183	28
Cronbach's Alpha: 0.865			

Table 72: Reliability and Descriptive Statistics for Naturalness (Extraversion Experiment)

Scale: Naturalness (Traits)

Mean	Variance	Std. Deviation	N of Items
15.1379	14.766	3.84266	5

Item Statistics

	Mean	Std. Deviation	N of cases
Natural	3.1379	1.15648	29
Humanlike	2.4483	1.02072	29
Conscious	3.0345	.98135	29
Lifelike	2.6897	.89056	29
Animate	3.8276	.92848	29
Cronbach's Alpha: 0.827			

9 REFERENCES

1. *The American Heritage® Dictionary of the English Language*. Fourth Edition ed: Houghton Mifflin Company. 2000.

2. *Merriam-Webster's Medical Dictionary*: Merriam-Webster, Inc. . 2002.

3. Frijda, N.H., *Varieties of Affect: Emotions and Episodes, Moods, and Sentiments*, in *The Nature of Emotion: Fundamental Questions*, P. Ekman, Davidson, R.J., Editor, Oxford University Press: New York. 1994

4. Watson, D., Clark, L.A., Tellegen, A., *Mood and Temperament*: The Guilford Press. 2000.

5. Matthews, G., *Cognitive Science Perspective on Personality and Emotion*. 3-30. 1997.

6. Revelle, W., *Personality Processes.* Annual Review of Psychology. **46**: p. 295-328, 1995.

7. Moffat, D., *Personality Parameters and Programs.* Creating Personalities for Synthetic Actors: p. 120-165, 1997.

8. Breckler, S.J., Wiggins, E.C., *On Defining Attitude and Attitude Theory: Once More with Feeling*, in *Attitude Structure and Function*, A.R. Pratkanis, Breckler, S.J., Greenwald, A.G. , Editor, Lawrence Erlbaum Associates. p. 407-429. 1989

9. Damasio, A.R., *Descartes' Error*, New York: G.P. Putnam's Sons. 1994.

10. Davidson, R.J., *On Emotion, Mood and Related Affective Constructs*, in *Nature of Emotion: Fundamental Questions*, D. Ekman P., R., Editor, Oxford Unviversity Press: New York. 1994

11. Frijda, N.H., *Emotions are Functional, Most of the Time*, in *The Nature of Emotions: Fundamental Questions*, D. Ekman P., R., Editor, Oxford University Press. 1994

12. Nass, C., Moon, Y., Morkes, J., Kim, E-Y., Fogg, B.J., *Computers are social actors: A review of current research*, in *Moral and ethical issues in human-computer interaction*, B. Friedman, Editor, CSLI Press: Stanford, CA. 1997

13. Oatley, K., Jenkins, J.M., *Understanding Emotions*, Cambridge, MA: Blackwell Publishers Inc. 1996.

14. Breazeal, C., Brooks, A., Gray, J., Hoffman, G., Kidd, C., Lee, H., Lieberman, J., Lockerd, A., Chilongo, D., *Tutelage and Collaboration for Humanoid Robots.* International Journal of Humanoid Robotics. **1**(4): p. 315-348, 2004.

15. Fong, T., Nourbakhsh, I., Dautenhahn, K., *A Survey of Socially Interactive Robots.* Robotics and Autonomous Systems. **42**: p. 143-166, 2003.

16. Goldberg, K., *Plenary Address.* Workshop on Performance Metrics for Intelligent Systems, 2010.

17. LeDoux, J., *The Emotional Brain*, New York, NY: Touchstone. 1996.

18. Reeves, B., Nass, C., *The Media Equation*, Cambridge: Cambridge University Press. 1996.

19. Ortony, A., Clore, G.L., Collins, A., *The Cognitive Structure of Emotions*, Cambridge: Cambridge University Press. 1988.

20. Picard, R.W., *Affective Computing*, The MIT Press: Cambridge, MA. 1997.

21. Brna, P., Cooper, B., *Marching to the Wrong Distant Drum: Pedagogic Agents, Emotion and Student Modelling.* International Conference on User Modeling, Workshop on Attitudes, Personality, and Emotions in User-Adapted Interaction, 2001.

22. Conati, C., Maclaren, H., *Evaluating a Probabilistic Model of Student Affect.* Intelligent Tutoring Systems: p. 55-66, 2004.

23. Lisetti, C., Nasoz, F., *MAUI: a multimodal affective user interface.* ACM Multimedia International Conference, 2002.

24. Johnson, M.P., Wilson, A., Kline, C., Blumberg, B., Bobick, A., *Sympathetic Interfaces: Using a Plush Toy to Direct Synthetic Characters.* CHI, 1999.

25. Paiva A., P.R., Chaves R., Vala M., Bullock A., Andersson G., Höök K., *Paiva A., Prada R., Chaves R., Vala M., Bullock A., Andersson G., Höök K.* ICMI, 2003.

26. Klein, J., Moon, Y., Picard, R.W., *This Computer Responds to User Frustration: Theory, Design, and Results. .* Interacting with Computers. **14**: p. 119-140, 2002.

27. Rani, P., Sarkar, N., Smith, C.A., *Affect-Sensitive Human-Robot Cooperation - Theory and Experiments.* International Conference on Robotics and Automation, 2003.

28. Ekman, P., *Moods, Emotions and Traits*, in *Nature of Emotion: Fundamental Questions*, D. Ekman P., R., Editor, Oxford University Press: New York. 1994

29. Fox, E., *Emotion Science: Cognitive and Neuroscientific Approaches to Understanding Human Emotions*: Palgrave MacMillan. 2008.

30. Lazarus, R.S., *The Stable and the Unstable in the Emotion*, in *The Nature of Emotion: Fundamental Questions*, P. Ekman, Davidson, R.J., Editor, Oxford University Press. 1994

31. Watson, D., Clark, L.A., *Emotions, Moods, Traits and Temperaments: Conceptual Distinctions and Empirical Findings*, in *The Nature of Emotion: Fundamental Questions*, D. Ekman P., R., Editor, Oxford Unviversity Press: New York. 1994

32. Kirby, R., Forlizzi, J., Simmons, R., *Affective Social Robots.* Robotics and Autonomous Systems. **58**: p. 322-332, 2010.

33. Miwa, H., Takanishi, A., Takanobu, H., *Experimental Study on Robot Personality for Humanoid Head Robot,* IROS: p. 1183-1188, 2001.

34. Miwa, H., Okuchi, T., Itoh, K., Takanobu, H., Takanishi, A., *A New Mental Model for Humanoid Robots for Human Friendly Communication.* ICRA, 2003.

35. Miwa, H., . et al., *Effective Emotional Expressions with Emotion Expression Humanoid Robot WE-4RII.* IROS, 2004.

36. Lisetti, C.L., Brown, S.M., Alvarez, K., Marpaung, A.H., *A Social Informatics Approach to Human–Robot Interaction With a Service Social Robot.* IEEE Transactions on Systems, Man and Cybernetics. **34**(2): p. 195-209, 2004.

37. Lisetti, C., Marpaung, A., *A three-layered architecture for socially intelligent agents: modeling the multilevel process of emotions*, in *1st International Conference on Affective Computing and Intelligent Interaction*. 2005: Beijing, China.

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

39. Murphy, R., Mali, A., *Lessons Learned in Integrating Sensing Into Autonomous Mobile Robot Architectures.* Journal of Experimental and Theoretical Artificial Intelligence. **9**(2): p. 191-220, 1997.

40. Fukuda, T., Jung, M., Nakashima, M., Arai, F., Hasegawa, Y., *Facial expressive robotic head system for human-robot communication and its application in home environment* Proceedings of the IEEE, 2004.

41. Ekman, P., Friesen, W.V., *Facial Action Coding System*, Palo Alto: Consulting Psychologist Press. 1978.

42. Leite, I., Pereira, A., Martinho, C., Paiva, A., *Are Emotional Robots More Fun to Play With?*, in *17th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*. 2008: Munich, Germany.

43. Castellano, G., Leite, I., Pereira, A., McOwan, P., *It's All in the Game: Towards an Affect Sensitive and Context Aware Game Companion*, in *Affective Computing and Intelligent Interaction and Workshop*. 2009: Amsterdam, Netherlands.

44. Wada, K., Shibata, T., Saito, T., Tanie, K., *Effects of Robot-Assisted Activity for Elderly People and Nurses at a Day Service Center.* Proceedings of the IEEE. **92**(11): p. 1780-1788, 2004.

45. Wada, K., Shibata, T., *Robot Therapy in a Care House - Change of Relationship among the Residents and Seal Robot during a 2-month Long Study*, in *16th IEEE International Conference on Robot & Human Interactive Communication*. 2007: Jeju, Korea.

46. Wada, K.I., Y. ; Inoue, K. ; Uehara, R., *Development and preliminary evaluation of a caregiver's manual for robot therapy using the therapeutic seal robot Paro*, in *19th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*. 2010: Viareggio, Italy.

47. Breazeal, C., *Emotion and Sociable Humanoid Robots.* International Journal of Human-Computer Studies. **59**: p. 119-155, 2003.

48. Arkin, R.C., Fujita, M., Takagi, T., Hasegawa, R., *An Ethological and Emotional Basis for Human-Robot Interaction.* Robotics and Autonomous Systems. **42**(3-4), 2003.

49. Esau, N., Kleinjohann, B., Kleinjohann, L., Stichling, D., *MEXI: machine with emotionally eXtended intelligence.* Design and application of hybrid intelligent systems: p. 961-970, 2003.

50. Canamero, L., *Playing the Emotion Game with Felix: What Can a Lego Robot Tell Us about Emotion?*, in *Socially Intelligent Agents*, K. Dautenhahn, Bond, A.H., Canamero, L., Edmonds, B., Editor, Kluwer Academic Publishers. 2002

51. Scheeff, M., Pinto, J., Rahardja, K., Snibbe, S., Tow, R., *Experiences with Sparky, a Social Robot*, in *Socially Intelligent Agents*, K. Dautenhahn, Bond, A.H., Canamero, L., Edmonds, B., Editor, Kluwer Academic Publishers. 2002

52. Ogata, T., Sugano, S., *Emotional Communication between Humans and the Autonomous Robot which has the Emotion Model.* ICRA: p. 3177-3182, 1999.

53. Schulte, J., Rosenberg, C., Thrun, S. , *Spontaneous, Short-term Interaction with Mobile Robots.* ICRA: p. 658-663, 1999.

54. Thrun, S.e.a., *MINERVA: A Second-Generation Museum Tour-Guide Robot.* ICRA, 1999.

55. Murray, J.C., Canamero, L. and Hiolle, A., , *Towards a Model of Emotion Expression in an Interactive Robot Head*, in *IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*. 2009.

56. Nakagawa, K.S., K.; Ishiguro, H.; Akimoto, T.; Hagita, N., *Motion modification method to control affective nuances for robots*, in *IEEE IROS*. 2009.

57. Wu, T., Butko, N.J., Ruvulo, P., Bartlett, M.S., and Movellan, J.R., *Learning to Make Facial Expressions*, in *IEEE International Conference on Development and Learning*. 2009.

58. Goris, K., Saldien, J., and Lefeber D., *Probo, a testbed for Human Robot Interaction*, in *ACM/IEEE International Conference on Human-Robot Interaction*. 2009.

59. Sosnowski, S., Bittermann, A., Kuhnlenz, K., and Buss, M., *Design and Evaluation of Emotion-Display EDDIE*, in *IEEE IROS*. 2006.

60. Yan, C., Peng, W., Lee, K.M., Jin, S., *Can robots have personality?An empirical study of personality manifestation, social responses, and social presence in human-robot interaction.* 54th Annual Conference of the International Communication Association, 2004.

61. Eklundth, K., Green, A., Huttenraunch, H., *Social and collaborative aspects of interaction with a service robot.* Robotics and Autonomous Systems, Special Issue on Socially Interactive Robots, 2003.

62. Sony. *Sony QRIO*. Available from: http://www.sony.net/SonyInfo/QRIO/top_nf.html.[accessed March 8, 2005]

63. Nagao, K., *Situated Conversation with a Communicative Interface Robot.* The First International Workshop on Intelligent Media Technology for Communicative Reality, 2002.

64. NEC. *NEC Personal Robot Center*. Available from: http://www.incx.nec.co.jp/robot/english/robotcenter_e.html; date accessed: March 8, 2005

65. Bethel, C.L., Murphy, R.R., *Review of Human Studies Methods in HRI and Recommendations.* Int Journal of Social Robotics. **2**: p. 347-359, 2010.

66. Bethel, C.L., Murphy, R. R., *Emotive Non-Anthropomorphic Perceived as More Calming, Friendly, and Attentive for Victim Management*, in *AAAI Fall Symposium: Dialog with Robots*. 2010: Arlington, VA.

67. Goetz, J., Kiesler, S. , *Cooperation with a robotic assistant.* ACM CHI'2002 Conference on Human Factors in Computing Systems: p. 578-579, 2002.

68. Mutlu, B., Osman, S., Forlizzi, J., Hodgins, J., Kiesler, S., *Task Structure and User Attributes as Elements of Human-Robot Interaction Design*, in *15th IEEE international workshop on robot and human interactive communication*. 2006.

69. Bruce, A., Nourbakhsh, I., Simmons, R., *The Role of Expressiveness and Attention in Human-Robot Interaction.* Proceedings of the IEEE International Conference on Robotics and Automation 2002.

70. Kahn, P.H., Friedman, B., Perez-Granados, D.R., Frier, N.G., *Robotic Pets in the Lives of Preschool Children.* CHI, 2004.

71. Friedman, B., Kahn, P.H., Hagman, J., *Hardware Companions? – What Online AIBO Discussion Forums Reveal about the Human-Robotic Relationship.* CHI, 2003.

72. Libin, A.V., Libin, E.V., *Person–Robot Interactions From the Robopsychologists' Point of View: The Robotic Psychology and Robotherapy Approach.* Proceedings of the IEEE, 2004.

73. Shibata, T., Tashima, T., Tanie, K., *Emergence of Emotional Behavior through Physical Interaction between Human and Robot.* ICRA, 1999.

74. Dautenhahn, K., Werry, I., Rae, J., Dickerson, P., Stribling, P., Ogden, B., *Robotic Playmates: Analysing Interactive Competencies of Children with Autism Playing with a Mobile Robot*, in *Socially Intelligent Agents*, K. Dautenhahn, Bond, A.H., Canamero, L., Edmonds, B., Editor, Kluwer Academic Publishers. 2002

75. Michaud, F., Theberge-Turmel, C., *Mobile Robotic Toys and Autism: Observations of Interaction*, in *Socially Intelligent Agents*, K. Dautenhahn, Bond, A.H., Canamero, L., Edmonds, B., Editor, Kluwer Academic Publishers. 2002

76. Dautenhahn, K., Werry, I., *A Quantitative Technique for Analysing Robot-Human Interactions.* International Conference on Intelligent Robots and Systems, 2002.

77. Psathas, G., *Conversation Analysis: The Study of Talk-In Interaction*, London: Sage. 1992.

78. Kanda, T., Ishiguro, H., Imai, M., Ono, T., *Development and Evaluation of Interactive Humanoid Robots.* Proceedings of the IEEE, 2004.

79. Scholtz, J., Bahrami, S., *Human-robot interaction: development of an evaluation methodology for the bystander role of interaction.* International conference on Systems, Man and Cybernetics, 2003.

80. Bartneck, C., Kulic, D., Croft, E., Zoghbi, S., *Measurement instruments for the anthropomorphism, animacy, likeability, perceived intelligence, and perceived safety of robots.* Int Journal of Social Robotics. **1**: p. 71-81, 2009.

81. Velásquez, J., *Modeling Emotion-Based Decision-Making.* AAAI Fall Symposium on Emotional and Intelligent: The Tangled Knot of Cognition: p. 164-169, 1998.

82. Izard, C.E., *Four Systems for Emotion Activation: Cognitive and Noncognitive Processes.* Psychological Review. **100**(1): p. 68-90, 1993.

83. El-Nasr, M.S., Yen, J., Ioerger, T., *FLAME - A Fuzzy Logic Adaptive Model of Emotions.* Autonomous Agents and Multi-agent Systems. **3**: p. 219-257, 2000.

84. Hashimoto, T., Hamada, T., Akazawa, T., *Proposal of Emotion-driven Control Model for Robot-Assisted Activity.* IEEE International Symposium On Computational Intelligence in Robotics and Automation, 2003.

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

86. Murphy, R.M., Lisetti, C., Tardif, R., Irish, L., Gage, A., *Emotion-based control of cooperating heterogeneous mobile robots.* IEEE Transactions on Robotics and Automation. **18**(5): p. 744-757, 2002.

87. Scheutz, *An Artificial Life Approach to the Study of Basic Emotions.* Annual Meeting of the Cognitive Science Society, 2004.

88. Arbib, M., *Schema Theory*, in *The Handbook of Brain Theory and Neural Networks*, S. Shapiro, Editor, MIT Press. p. 830-833. 1992

89. Lee-Johnson, C.P., Carnegie, D.A., *Emotion-based Parameter Modulation for a Hierarchical Mobile Robot Planning and Control Architecture*, in *IEEE IROS*. 2007.

90. Domínguez, C., Hassan, H., Crespo, A., *Behavioural Affective Modulation of Service Robotic Agents*, in *IEEE Annual Conference on Industrial Electronics*. 2006.

91. Avila-Garcia, O., Canamero, L., *Using Hormonal Feedback to Modulate Action Selection in a Competitive Scenario.* From Animals to Animats: Proc. 8th Intl. Conf. on Simulation of Adaptive Behavior SAB'04), 2004.

92. Michaud, F., Robichaud, E., Audet, J., *Using motives and artificial emotions for long-term activity of an autonomous robot.* International Conference on Autonomous Agents 2001.

93. Gadanho, S.C., *Learning behavior-selection by emotions and cognition in a multi-goal robot task.* The Journal of Machine Learning Research 2003.

94. Malfaz, M., Salichs, M.A., *A new architecture for autonomous robots based on emotions.* Fifth IFAC Symposium on Intelligent Autonomous Vehicles, 2004.

95. Mateas, M., *An Oz-Centric Review of Interactive Drama and Believable Agents*, in *Technical Report CMU-CS-97-156*. 1997, Carnegy Melon University, School of Computer Science

96. El Jed, M., Pallamin, N., Dugdale, J., Parvard, B., *Modelling Character Emotion in an Interactive Virtual Environment.* AISB 2004 Convention, Symposium on Language, Speech and Gesture for Expressive Characters, 2004.

97. Egges, A., Kshirsagar, S., Magnenat-Thalmann, N., *Generic Personality and Emotion Simulation for Conversational Agents.* Computer Animation and Virtual Worlds. **15**(1): p. 1-13, 2004.

98. Prendinger, H., Descamps, S., Ishizuka, M., *Scripting affective communication with life-like characters in web-based interaction systems.* Applied Artificial Intelligence. **16**(7-8): p. 519-553, 2002.

99. Bates, J., *The Role of Emotion in Believable Agents.* Communications of the ACM, Special Issue on Agents. **37**(7): p. 122-125, 1994.

100. Elliott, C., Siegle, G. , *Variables Influencing the Intensity of Simulated Affective States.* AAAI Spring Symposium on Reasoning about Mental States: Formal Theories and Applications: p. 58-67, 1993.

101. Gratch, J., Marsella, S., *A Domain-independent Framework for Modeling Emotion.* Journal of Cognitive Systems Research. **5**(4): p. 269-306, 2004.

102. Marsella, S., Gratch, J., *EMA: A Model of Emotional Dynamics.* Journal of Cognitive Systems Research. **10**(1): p. 70-90, 2009.

103. de Rosis, F., Pelachaud, C., Poggi, E., Carofiglio, V., de Carolis, B., *From Greta's Mind to her Face: Modelling the Dynamics of Affective States in a Conversational Embodied Agent.* International Journal of Human-Computer Studies. **59**(1-2): p. 81-118, 2003.

104. Rizzo, P., *Realizing Believable Agents: An Integration of the "Author-based" and the "Model-based" Approaches.* AI Communications, Special Issue on "Animated Interface Agents". **13**: p. 145-168, 2000.

105. Ball, G., Breese, J., ed. *Emotion and Personality in a Conversational Agents*. Embodied Conversational Agents, ed. J. Cassell, Sullivan, J., Prevost, S., Churchill, E. 2000, The MIT Press: Cambridge, MA.

106. Tomlinson, B., Downie, M., Berlin, M., Gray, J., Lyons, D., Cochran, J., Blumberg, B., *Leashing the AlphaWolves: mixing user direction with autonomous emotion in a pack of semi-autonomous virtual characters.* ACM SIGGRAPH/Eurographics symposium on Computer animation p. 7-14, 2002.

107. Moshkina, L., Arkin, R.C. *Human Perspective on Affective Robotic Behavior: A Longitudinal Study*. in *IROS*. Edmonton, Canada: IEEE. 2005.

108. Moshkina, L., Arkin, R.C., *On TAMEing Robots.* Interanational Conference on Systems, Man and Cybernetics, 2003.

109. Arkin, R., *Motor-Schema based Mobile Robot Navigation.* International Journal of Robotics Research. **8**(4): p. 92-112, 1989.

110. Moshkina, L., Arkin, R.C., Lee, J., and Jung, H., *Time Varying Affective Response for Humanoid Robots*, in *International Conference on Social Robotics (ICSR)*. 2009: Seoul, KR.

111. Clore, G.L., Wyer, R.S.Jr., Dienes, B., Gasper, K., Gohm, C, Isbell, L., *Affective Feelings as Feedback: Some Cognitive Consequences*, in *Theories of Mood and Cognition: a User's Guidebook*, L.L. Martin, Clore, G.L., Editor, Lawrence Erlbaum Associates, Publishers: Mahwah, NJ. 2001

112. Buss, D.M., *Human Nature and Individual Differences: The Evolution of Human Personality*, in *Handbook of Personality: Theory and Research*, L.A. Pervin, John, O.P., Editor, The Guilford Press: New York. 1999

113. Matthews, G., Derryberry, D. , *Personality and Emotion: Cognitive Science Perspectives.* Advances in Personality Psychology. **1**: p. 199-237, 2000.

114. McCrae, R.R., Costa, P.T. , *Toward a new generation of personality theories: theoretical contexts for the five-factor model.* Five-Factor Model of Personality: p. 51-87, 1996.

115. Goldberg, L.R., *An Alternative "Description of Personality": The Big Five Factor Structure.* Journal of Personality & Social Psychology. **59**: p. 1216-1229, 1990.

116. John, O.P., Srivastava, S., *The Big Five Trait Taxonomy: History, Measurement, and Theoretical Perspectives*, in *Handbook of Personality: Theory and Research*, L.A. Pervin, John, O.P., Editor, The Guilford Press: New York. 1999

117. Gosling, S.D., John, O.P., *Personality Dimensions in Nonhuman Animals: A Cross-Species Review.* Current Directions in Psychological Science: p. 69-75, 1999.

118. Costa, P.T., McCrae, R.R., *NEO PI-R Professional Manual.* Psychological Assessment Resources, 1992.

119. Elliot, A.J., Thrash T.M. , *Approach-Avoidance Motivation in Personality: Approach and Avoidance Temperaments and Goals.* Journal of Personality and Social Psychology. **82**(5): p. 804-818, 2000.

120. Panksepp, J., *Emotions As Natural Kinds within the Mammalian Brain*, in *Handbook of Emotions*, M. Lewis, Haviland-Jones, J.M., Editor, The Guilford Press: New York. 2000

121. Izard, C.E., Ackerman, B.P., *Motivational, Organizational, and Regulatory Functions of Discrete Emotions*, in *Handbook of Emotions*, M. Lewis, Haviland-Jones, J.M., Editor, The Guilford Press: New York. 2000

122. Johnson-Laird, P.N., Oatley, K., *Cognitive and Social Construction in Emotions*, in *Handbook of Emotions*, M. Lewis, Haviland-Jones, J.M., Editor, The Guilford Press: New York. 2000

123. Scherer, K.R., *Appraisal Considered as a Process of Multilevel Sequential Checking*, in *Appraisal Processes in Emotion*, K.R. Scherer, Schorr, A., Johnstone, T., Editor, Oxford Universtiy Press: Oxford. 2001

124. Smith, C.A., Kirby, L.D., *Toward Delivering on the Promise of Appraisal Theory*, in *Appraisal Processes in Emotion*, K.R. Scherer, Schorr, A., Johnstone, T., Editor, Oxford Universtiy Press: Oxford. 2001

125. Keltner, D., Ekman, P., *Facial Expression of Emotion*, in *Handbook of Emotions*, M. Lewis, Haviland-Jones, J.M., Editor, The Guilford Press: New York. 2000

126. Canamero, L.D., *Designing Emotions for Activity Selection in Autonomous Agents*, in *Emotions in Humans and Artifacts*, R. Trappl, Petta, P., Payr, S., Editor, The MIT Press. 2003

127. Rolls, E.T., *A Theory of Emotion, Its Function, and Its Adaptive Value*, in *Emotions in Humans and Artifacts*, R. Trappl, Petta, P., Payr, S., Editor, The MIT Press. 2003

128. Lazarus, R.S., *Relational Meaning and Discrete Emotions*, in *Appraisal Processes in Emotion*, K.R. Scherer, Schorr, A., Johnstone, T., Editor, Oxford Universtiy Press: Oxford. 2001

129. Frijda, N.H., *The Psychologists' Point of View*, in *Handbook of Emotions*, M. Lewis, Haviland-Jones, J.M., Editor, The Guilford Press: New York. 2000

130. Lemerise, E.A., Dodge, K.A., *The Development of Anger and Hostile Interactions*, in *Handbook of Emotions*, M. Lewis, Haviland-Jones, J.M., Editor, The Guilford Press: New York. 2000

131. Rozin, P., Haidt, J., McCauley, C.R., *Disgust: the Body and Soul Emotion*, in *Handbook of Cognition and Emotion*, T. Dalgleish, Power, M.J., Editor. 1999

132. Ortony, A., *On Making Believable Emotional Agents Believable*, in *Emotions in Humans and Artifacts*, R. Trappl, Petta, P., Payr, S., Editor, The MIT Press. 2003

133. Davidson, R.J., *Neuropsychological Perspectives on Affective Styles and Their Cognitive Consequences*, in *Handbook of Cognition and Emotion*, T. Dalgleish, Power, M.J., Editor, John Wiley and Sons. p. 103-124. 1999

134. McFarland, D.J., Bosser, T., *Intelligent Behavior in Animals and Robots.* 1993.

135. Arkin, R., Fujita, M., Takagi, T., Hasegawa, R. , *Ethological Modeling and Architecture for an Entertainment Robot*, in *ICRA*. 2001. p. 453-458.

136. Breazeal, C., *Designing Sociable Robots*: The MIT Press. 2002.

137. Velasquez, J.D., *An Emotion-Based Approach to Robotics.* IROS: p. 235-240, 1999.

138. Bless, H., *Mood and the Use of General Knowledge Structures*, in *Nature of Emotion: Fundamental Questions*, D. Ekman P., R., Editor, Oxford Unviversity Press: New York. 1994

139. Isen, A.M., *Positive Affect*, in *Handbook of Cognition and Emotion*, T. Dalgleish, Power, M.J., Editor, John Wiley and Sons. 1999

140. Thayer, R.E., *The Origin of Everyday Moods*: Oxford University Press. 1996.

141. Schwarz, N., Clore, G.L., *Mood, Misattribution, and Judgments of Well-Being: Informative and Directive Functions of Affective States.* Journal of Personality & Social Psychology. **45**: p. 513-523, 1983.

142. Clore, G.L., Martin, L.L., *Introduction*, in *Theories of Mood and Cognition: a User's Guidebook*, L.L. Martin, Clore, G.L., Editor, Lawrence Erlbaum Associates, Publishers: Mahwah, NJ. 2001

143. Schwarz, N., *Feelings as Information: Implications for Affective Influences on Information Processing*, in *Theories of Mood and Cognition: a User's Guidebook*, L.L. Martin, Clore, G.L., Editor, Lawrence Erlbaum Associates, Publishers: Mahwah, NJ. 2001

144. Bless, H., *Mood and the Use of General Knowledge Structures*, in *Theories of Mood and Cognition: a User's Guidebook*, L.L. Martin, Clore, G.L., Editor, Lawrence Erlbaum Associates, Publishers: Mahwah, NJ. 2001

145. Forgas, J.P., *The Affect Infusion Model (AIM): An Integrative Theory of Mood Effects on Cognition and Judgments*, in *Theories of Mood and Cognition: a User's Guidebook*, L.L. Martin, Clore, G.L., Editor, Lawrence Erlbaum Associates, Publishers: Mahwah, NJ. 2001

146. Nass, C., Jonsson, I.-M.,Harris, H.,Reaves, B.,Endo, J.,Brave, S. & Takayama, L., *Improving Automotive Safety by Pairing Driver Emotion and Car Voice Emotion*, in *SIGCHI Conference on Human factors in Computing Systems*. 2005: Portland, OR.

147. Nass, C.I.M., Y., *Machines and mindlessness: Social responses to computers.* Journal of Social Issues. **56**(1): p. 81-103, 2000.

148. Park, S., Moshkina, L., and Arkin, R.C. *Mood as an Affective Component for Robotic Behavior with Continuous Adaptation via Learning Momentum*. in *Humanoids*. Nashville, TN. 2010.

149. Olson, J.M., Zanna, M.P., *Attitudes and Attitude Change.* Annual Review of Psychology. **44**(117-54), 1993.

150. Ajzen, I., *Attitude Structure and Behavior*, in *Attitude Structure and Function*, A.R. Pratkanis, Breckler, S.J., Greenwald, A.G., Editor, Lawrence Erlbaum Associates. 1989

151. Fabrigar, M., Wegener, *Structure of Attitudes.*

152. Snyder, M., DeBono, K.G., *Understanding the Functions of Attitudes: Lessons from Personality and Social Behavior*, in *Attitude Structure and Function*, A.R. Pratkanis, Breckler, S.J., Greenwald, A.G., Editor, Lawrence Erlbaum Associates. 1989

153. Greenwald, A.G., *Why Are Attitudes Important*, in *Attitude Structure and Function*, A.R. Pratkanis, Breckler, S.J., Greenwald, A.G., Editor, Lawrence Erlbaum Associates. 1989

154. Clore, G.C., *Why Emotions are Felt*, in *The Nature of Emotion: Fundamental Questions*, P. Ekman, Davidson, R.J., Editor, Oxford University Press. 1994

155. Aamodt, A., Plaza, E., *Case-Based Reasoning: Foundational Issues, Methodological Variations, and System Approaches.* AI Communications. **7**(1): p. 39-59, 1994.

156. Kira, Z., and Arkin, R.C., *Forgetting Bad Behavior: Memory Management for Case-Based Navigation*, in *International Conference on Intelligent Robots and Systems (IROS)*. 2004. p. 3145-3152.

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

158. Lippa, R., *The Nonverbal Display and Judgement of Extraversion, Masculinity, Femininity, and Gender Diagnosticity: A Lens Model Analysis.* Journal of Research in Personality. **32**: p. 80-107, 1998.

159. Riggio, E.R., Lippa, R., Salinas, C., *The Display of Personality in Expressive Movement.* Journal of Personality Research. **24**: p. 16-31, 1990.

160. Buytendijk, F., *The Mind of the Dog*, Arno Press: New York. 1973.

161. Fox, M., *The Dog: Its Domestication and Behavior*, Garland: New York. 1978.

162. Watson, D., Clark, L.A., Tellegen, A., *Development and validation of brief measures of positive and negative affect: The PANAS scales.* Journal of Personality & Social Psychology. **56**(6): p. 1063-1070, 1998.

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

164. Hedrick, T.E., Bickman, L., Rog, D.J., *Applied Research Design*: Sage Publications. 1993.

165. Park, S., Moshkina, L., and Arkin, R.C., *Recognizing Nonverbal Affective Behavior in Humanoid Robots*, in *Proc. of 11th Intelligent Autonomous Systems Conference*. 2010: Ottawa, Canada.

166. Moshkina, L., Park, S., Arkin, R.C., Lee, J., and Jung, H., *TAME: Time Varying Affective Response for Humanoid Robots.* International Journal of Social Robotics, 2011.

167. Park, S., Moshkina, L., and Arkin, R.C., *Mood as an Affective Component for Robotic Behavior with Continuous Adaptation via Learning Momentum*, in *Humanoids*. 2010: Nashville, USA.

168. Brooks, A.a.A., R.C., *Behavioral Overlays for Non-Verbal Communication Expression on a Humanoid Robot.* Autonomous Robots. **22**(1): p. 55-75, 2007.

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

170. Arkin, R., Balch, T. , *AuRA: Principles and Practice in Review.* Journal of Experimental and Theoretical Artificial Intelligence. **9**(2): p. 175-189, 1997.

171. Arkin, R.C., *Behavior-based Robotics*: MIT Press. 1998.

172. Bethel, C.L., Murphy, R.R., *Survey of Non-facial/Non-verbal Affective Expressions for Appearance-Constrained Robots.* IEEE Transactions on Systems, Man and Cybernetics. **38**(1): p. 83-92, 2008.

173. Argyle, M., *Bodily Communication*, London, U.K: Methuen. 1975.

174. Nass, C., Lee, K.M., *Does Computer-Synthesized Speech Manifest Personality? Experimental tests of recognition, similarity-attraction, and consistency-attraction.* Journal of Experimental Psychology: Applied. **7**(3): p. 171-181, 2001.

175. Lee, K.M., Peng, W., Jin, S.A., and Yan, C., *Can robots manifest personality? An empirical test of personality recognition, social responses, and social presence in human-robot interaction.* Journal of Communication. **56**(4): p. 754-772, 2006.

176. Gallaher, P.E., *Individual differences in nonverbal behavior: Dimensions of style.* Journal of Personality and Social Psychology. **63**: p. 133–145, 1992.

177. Gifford, R., *Mapping nonverbal behavior on the interpersonal circle.* Journal of Personality & Social Psychology. **61**: p. 279–288, 1991.

178. *Center for Spoken Language Understanding (CSLU) toolkit*, in *Oregon Graduate Institute*, available at http://cslu.cse.ogi.edu; date accessed: June 29, 2010

179. Atkinson, A.P., Dittrich, W.H., Gemmell, A.J., Young, A.W., *Emotion perception from dynamic and static body expressions in point-light and full-light displays.* Perception. **33**(6): p. 717–746, 2004.

180. Wallbott, H.G., *Bodily expression of emotion.* European Journal of Social Psychology. **28**: p. 879–896, 1998.

181. Coulson, M., *Expressing emotion through body movement: A component process approach*, in *Animating Expressive Characters for Social Interactions*, R. Aylett, Cañamero, L., Editor, SSAISB Press. 2002

182. Mehrabian, A., Ksionsky, R., *A theory of Affiliation*, Lexington, MA: Heath. 1974.

183. Carney, D.R., Colvin, C.R., and Hall, J.A., *A thin slice perspective on the accuracy of first impressions.* Journal of Research in Personality **41**: p. 1054-1072, 2007.

184. Collier, G., *Emotional Expression*, Hillsdale, NJ: Lawrence Erlbaum Associates. 1985.

185. Judge, T.A., Bono, J.E., Ilies, R., Gerhardt, M.W., *Personality and leadership: A qualitative and quantitative review.* Journal of Applied Psychology. **87**(4): p. 765-780, 2002.

186. Radmacher, S.A., Martin, D.J., *Identifying Significant Predictors of Student Evaluations of Faculty Through Hierarchical Regression Analysis.* Journal of Psychology: Interdisciplinary and Applied **135**(3): p. 259-268, 2001.

187. Rushton, J.P., Murray, H.G., Erdle, S., *Combining Trait Consistency and Learning Specificity Approaches to Personality, with Illustrative Data on Faculty Teaching Performance.* Journal of Personality and Individual Differences. **8**(1): p. 59-66, 1987.

188. Damos, D.L., ed. *Multiple Task Performance*. 1991, Taylor & Francis: London.

189. Hart, S.G., Staveland, L.E., *Development of NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research*, in *Human Mental Workload*, P.A. Hancock, Meshkati, N., Editor, North Holland Press: Amsterdam. p. 239-250. 1988

190. Morris, J.D., *Observations: SAM: The Self-Assessment Manikin, An Efficient Cross-Cultural Measurement of Emotional Response.* Journal of Advertising Research, 1995.

191. Lang, P.J., Bradley, M.M., and Cuthbert, B.N., *International Affective Picture System (IAPS): Technical Manual and Affective Ratings*, Gainesville: University of Florida, Center for Research in Psychophysiology. 1999.

192. Messick, S., Jackson, D.N., *Acquiescence and the factorial interpretation of the MMPI.* Psychological Bulletin. **58**(4): p. 299-304, 1961.

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

194. Rani, P., Sarkar, N., Smith, C.A., Kirby, L.D., *Anxiety detecting robotic system—towards implicit human-robot collaboration.* Robotica. **22**(1): p. 85-95, 2004.

195. Scheutz, M.S., P.; Kramer, J.; and Middendorff, C., *The utility of affect expression in natural language interactions in joint human-robot tasks*, in *ACM International Conference on Human-Robot Interaction*: Utah, USA.

196. Goetz, J., Kiesler, S., & Powers, A., *Matching robot appearance and behavior to tasks to improve human-robot cooperation*, in *IEEE International Workshop on Robot and Human Interactive Communication (RO-MAN)*. 2003: Milbrae, CA, USA.

197. Palmer, J.K., & Loveland, J.M., *Further Investigation of the Psychometric Properties of Saucier's Big Five "Mini-Markers:" Evidence for Criterion and Construct Validity.* Individual Differences Research. **2**(3): p. 231-238, 2004.

198. Patrick, C.J., Lavoro, S.A., *Ratings of Emotional Response to Pictorial Stimuli: Positive and Negative Affect Dimensions.* Motivation and Emotion. **21**: p. 297-322, 1997.

199. Goldberg, L.R., *The Development of Markers for the Big-Five Factor Structure.* Psychological Assessment. **4**: p. 26-42, 1992.

200. Nunnally, J.C., *Psychometric theory*, New York: McGraw-Hill. 1978.

201. Osgood, C.E., Suci, G.J., Tannenbaum, P.H., *The Measurements of Meaning*, Champaign: University of Illinois Press. 1957.

202. Al-Hindawe, J., *Considerations when constructing a semantic differential scale.* La Trobe Papers in Linguistics. **9**, 1996.

203. Donmez, B., Pina, P. E., Cummings, M. L., *Evaluation criteria for human-automation performance metrics*, in *Performance Evaluation and Benchmarking of Intelligent Systems*, R. Madhavan, Editor, Springer. p. 21-40. 2009

204. Jung, H., Moshkina, L., Choi, H., Lee, J., and Arkin, R., *Reusable Affective Software Architecture for Humanoid Robots and Virtual Agents*, in *International Conference on Intelligent Virtual Agents (poster)*. 2010: Philadelphia, USA.