The Interplay of Context and Emotion for Non-Anthropomorphic Robots

Keith McGreggor[1], Jenay M. Beer[2], Bryan Wiltgen[1], Karl Jiang[1], & Andrea Thomaz[1]
[1]School of Interactive Computing & [2]School of Psychology
Georgia Institute of Technology
Atlanta, Georgia, USA
{keith.mcgreggor, jbeer, bryan.wiltgen, k.jiang}@gatech.edu, athomaz@cc.gatech.edu

***Abstract*—Household robots are becoming commonplace. The application of social cues, such as emotion, has the potential to make such robots easier to use and understand. However, it remains unclear how household robots can or should display emotion, and what considerations should be given to emotive behavior regarding the expected set of contexts in which the robot will operate. In this paper, we report the results of our systematic evaluation of the role both context and emotions play in the interpretation of context and emotion recognition of a non-anthropomorphic robot, the iRobot Roomba. Considerations, implications, and future work are discussed.**

Keywords: emotion; context; appraisal; social robotics

I. Introduction

Household robots are becoming more commonplace in the world [10]. Despite this increase in occurrence and acceptance, a persistent problem with household robots is the level of expertise needed to interact effectively with them, especially when the robot is in a state of requiring some assistance or support. Therefore, when designing more capable, intelligent robots, considerations must extend beyond simply improving the underlying technologies. To foster improved human-robot interaction, the robot designer must understand and pay particular heed to the social characteristics of robots.

It is generally accepted in the research community that people are willing to apply social characteristics, unbidden or otherwise, to technology and to technological artifacts. Humans have been shown to apply social characteristics to computers, even treating computers as teammates or as possessing personality, similar to human-human interaction [1]-[3]. A survey of recent iRobot Roomba™-specific human-robot interaction (HRI) literature suggests that there are at least three areas of research concentration regarding emotional designation:

- Without modification, the robot may be perceived by its owners as having affective traits [4][5][6];
- Users develop emotional connections to their robot, saying that they "love it" or considering it "part of the family" [7]; and
- Perception of the robot appears to change over time as interactions become more routine, wherein the robot begins to be perceived of more as an appliance and less as a pet [8].

Effective social interaction requires more than the application of social characteristics to a robot; however, as robotic technology becomes more advanced, robots are becoming more capable of actively demonstrating social cues. If applied to robots specifically designed to interact with untrained users, social cues may improve communication and facilitate natural social interaction [9]. However, such social cues must be chosen with care to ensure the intended human-robot interaction. Behaviors selected for demonstration by robots might prove useful in certain contexts while wholly inappropriate in others.

The overall goal of our research is to improve the emotional intelligence and emotional evocativeness of a robot, to facilitate a more optimal human-robot interaction, and therefore to require less skilled effort from the robot's owner or user in everyday interactions. We are focusing our research efforts on one such commonplace household robot, the Roomba. This is an inexpensive household vacuum robot that presents unique challenges concerning the demonstration of social cues, such as emotion. The Roomba has a non-humanoid design, very limited motor and visual systems, and a primitive audio system. We have chosen the Roomba because we theorize that there are a number of situations where the owner's understanding of the robot's present context could be improved by affording to the Roomba an extensive emotional repertoire. This repertoire will be one achieved programmatically by way of deliberately chosen actions rather than unintentional or ambiguous assignment due to the robot's intrinsic design.

In this work we find that task context has a significant effect upon a person's ability to accurately interpret a robot's selected emotion expression. In particular, we illustrate that an emotional display is most accurately interpreted when displayed in an appropriate task context, as compared to an incongruent context or without context altogether. This has important implications for the design and evaluation of emotional robots particularly non-anthropomoric emotional robots.

II. The Interplay of Context and Emotion

Before a robot's emotional repertoire can be developed, information about the relationship between emotion recognition and contexts must be discovered. Without knowing this relationship, one may inadvertently develop emotive behaviors that are interpreted as desired in one context but are interpreted entirely differently within another. In addition, deliberately but incorrectly chosen behaviors might cause users to interpret a particular context incorrectly.

We seek to know whether emotion recognition is influenced by contextual information, or the opposite: is context recognition influenced by emotional information? What is the nature of the interplay between contextual and emotional information? Are the two aspects mutually dependent? While it is generally understood that context and emotion are related [11], the authors wish to investigate this relationship as it relates specifically to HRI.

A. Significant considerations

We note that any robot itself is inseparably a part of the overall context as perceived by its owner or user. Any intentional or unintentional emotional evocativeness proscribed by either the robot's appearance or by the robot's presence within the context must be taken into account. Thus, the findings for a particular robot (or class of robots) will be idiosyncratic.

Equally important is the notion that the robot may have a limited capability for expression or range of mobility. While there exist robots on the market with more human-like features and more robust output capabilities (e.g., speech), we specifically chose to study the Roomba due to its limited intrinsic repertoire and non-anthropomorphic form-factor. We feel the Roomba's limited capabilities force a more deliberate approach to behavior selection, as it is significantly harder to ground the behaviors in human-human interactions than it would be if, for example, the Roomba had a human-like face.

Any purposefully constructed robot will have a finite task context within which there will be specific states it may wish to communicate to observers. Actions selected for those communications may elicit unintentional emotional recognition. We believe that the more non-humanoid the robot is in either design or behavior (or both), the more likely the misidentification of emotional intent. The entire repertoire of communicative behavior for such a robot must be designed with a notion of how the robot's present situation or context might cause those selected behaviors to be interpreted emotionally.

B. Research Questions

Within the domain of human-robot interaction, we wish to address two specific questions:

- How can the apparent context within which a person observes a robot influence emotion recognition, and in particular, can that contextual information influence a person's identification of an emotion?
- Similar to the previous question, how does a displayed emotive behavior effect a person's identification of the context in which that behavior occurs?

If the displayed emotion and context are incongruent with respect to context-emotion expectations, which aspect can be said to be more influential in emotion recognition (e.g., if a robot demonstrates a "negative" emotion during a "positive" context, will a person be able to identify that emotion correctly)?

We suggest that the interplay between emotion and context is such that incongruent pairings of emotion-context will result in less accurate emotion recognition, and similarly, that incongruent emotion-context will result in less accurate context recognition. We have designed and conducted an experiment that seeks to confirm this hypothesis. Using our methodology, we seek to demonstrate that the relationship between emotion and context is bidirectional and mutually enhancing, and that the bidirectional nature of that relationship plays a necessary and significant role in recognition of both context and emotion.

III. METHOD

A. Participants

The participants were 20 young adults (3 female and 17 male), between 18 and 30 years of age. The participants were undergraduate and graduate students from the Georgia Institute of Technology. The participants volunteered, and were not compensated for their time.

B. Materials

Video clips of an iRobot Roomba vacuum cleaner pet series were presented to participants using a secured website, and participants made responses using a standard mouse or touchpad.

The secured website that facilitated the experiment was custom built for this study (see Fig. 1). We chose this route over using a prebuilt solution because having an in-house experimental platform allowed for very fine-grained manipulation of the presentation of the video clips, such as using AJAX and Adobe Flash™ video to provide seamless transitions between video clips and questions. Additionally, having the entire code base for the site allows us to easily replicate and/or extend the experiment in the future.

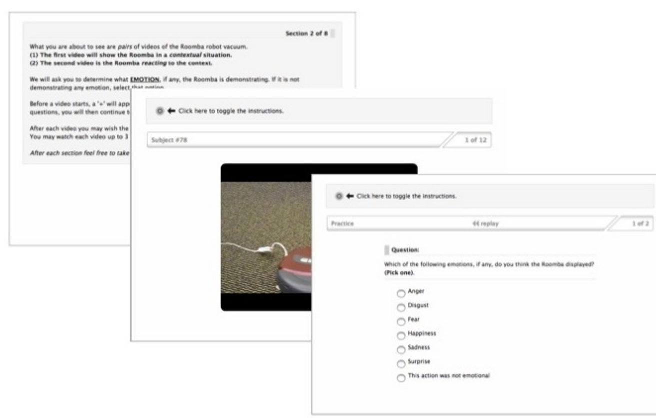

Figure 1. Secured experimental website.

C. Stimuli

Using an iRobot Roomba vacuum cleaner pet series, the experimenters identified distinct behaviors and contexts the robot may engage in. The behaviors were broadly categorized as movements, sounds, or lights (e.g., rotation, beeps, blinking, etc). While contexts were identified as situations where either something was wrong or the Roomba needed human intervention (e.g., the Roomba becomes stuck). Short video clips were created of the Roomba demonstrating each of these contexts (range of 4 to 6 s, M = 5.0 s) and each of these emotions (range of 3 to 10 s, M = 6.2 s), with the context preceding the emotion. The videos were edited so that the time between the beginning of the video and the beginning of the action (be it emotive behavior or context-related) and the time between the end of the video and the end of the action was as short as possible.

Several iterations of pilot testing were conducted where participants were asked to identify the context, as well as if emotion was present in each of the behaviors. The pilot data revealed 5 emotive behaviors and 6 contexts that were easily recognizable. See Table 1 for a description of each of these videos. Additionally, a pilot questionnaire was administered which asked the participants to identify emotion(s) that they would expect the Roomba to demonstrate in response to possible contexts.

TABLE I. DESCRIPTION OF EMOTION AND CONTEXT VIDEOS

Context Videos	
Context	***Description***
Bump	Roomba bumps into an object
Dirt	Roomba senses dirt on the ground
Dock (unsuccessful)	Roomba unsuccessfully docks
Edge	Roomba approaches an edge
Lift	Roomba is picked up by user
Stuck	Roomba becomes stuck between two objects
Emotion Videos	
Emotion	***Description***
Anger	Moving forward in a staccato manner while blinking a red light
Fear	Moving backward at a slow speed
Happiness	Display a constant blue light while producing a chirping, uptempo sound
Sadness	Moving forward at a slow speed while producing a low vibrato sound
Surprise	Starting forward, then moving backward at a fast speed while producing a rapid high-pitched beeping sound

The emotive behaviors and contexts were matched as congruent pairs and incongruent pairs (see Table 2). The congruent pairs were determined by the pilot questionnaire, and represent the context-emotion pairs that participants reported they would expect the Roomba to engage in. The experimenters then determined the incongruent pairs by selecting an emotion opposite of the congruent pair. The opposite emotion was chosen based upon its opposite arousal, stance or valence of the congruent emotion [12].

TABLE II. CONTEXT-EMOTION PAIRINGS

Congruent Videos		**Incongruent Videos**	
Context	***Emotion***	***Context***	***Emotion***
Bump	Surprise	Bump	Anger
Dirt	Happy	Dirt	Sad
Dock (unsuccessful)	Anger	Dock (unsuccessful)	Happy
Edge	Fear	Edge	Anger
Lift	Surprise	Lift	Sad
Stuck	Anger	Stuck	Happy

D. *Design*

A within subjects design was conducted. Each participant viewed Congruent Context-Emotion pairs, Incongruent Context-Emotion pairs, and a Control condition where they viewed each emotion and context video in isolation. Dependent measures were Recognition Accuracy for making an emotion identification response, and Recognition Accuracy for making a context identification response.

E. *Procedure*

Participants were first provided informed consent, which outlined the general aspects of the study as well as their rights as participants. After completion of the informed consent forms, the participants were provided with a description of how the web-based experiment worked and instructions about the practice and the experimental task.

During the practice session, the participants were presented with unrelated videos of a robot performing an action. The participants were instructed to watch the video. Then, when prompted, they were asked to select which emotion (or no emotion) they believed the robot demonstrated. The practice session was designed to allow each participant to become familiar with the web-based experiment.

After completion of the practice trials, the participants began the experimental session. They were presented with the stimuli, and then asked to either identify the emotion or the context. When asked to identify the emotion, participants could choose from the six basic emotions (anger, disgust, fear, happiness, sadness, and surprise) [13] as well as neutral. When asked to identify the context, participants were asked to choose from a list of possible contexts, which were derived from pilot testing.

The stimuli were presented in blocks. Three types of blocks were used:

- Congruent and incongruent pairs were presented and participants were asked to identify the emotion the Roomba demonstrated, if any;
- Congruent and incongruent pairs were presented and participants were asked to identify the context; and
- Emotive behavior and context videos were shown in isolation (non-pairs) and participants were asked to identify the emotion or context respectively.

Within each block, the trials were presented randomly. The order of blocks was counterbalanced to reduce carry-over effects. Each participant viewed six blocks, two of each type.

Immediately following the experimental session, the participants were asked to answer the web-based questionnaire, which asked about prior robot experience. Upon completion of the questionnaire, the participants were thanked for their time, and debriefed.

F. Hypotheses

Through careful pilot testing, the experimenters identified videos of task contexts and emotions that are readily identified. The experimenters systematically created three conditions: congruent context-emotion pairings, incongruent context-emotion pairings, and a control (task context videos only, or emotion videos only). On the basis of the pilot data, the experimenters did not expect to find a difference between the congruent video pairings and the control, for either emotion recognition or context recognition. However, the incongruent context-emotion pairings were expected to influence recognition accuracy, generating the two following hypotheses:

[H1]: Emotion Recognition will be influenced by context. More specifically, an emotion displayed in an incongruent task context will be more difficult to accurately identify than one displayed either by itself (without context) or in a congruent task context.

[H2]: Task Context Recognition will be influenced by emotion. More specifically, a task context displayed in an incongruent emotion will be more difficult to accurately identify than one displayed either by itself (without emotion) or in a congruent emotion.

IV. Results

Separate repeated measures analyses of variance (ANOVA) were conducted to test the effects of emotion-context congruency on emotion recognition and context recognition accuracy. Bonferroni corrections were applied where appropriate. Very infrequently a video failed to display properly. In these cases, we chose to remove those instances from our analysis.

A. Emotion Recognition

A one-way repeated measures ANOVA was carried out to determine if emotion recognition accuracy differed as a function of context-emotion congruency. A statistically significant main effect was found, indicating that the emotion recognition accuracy mean scores differed significantly $F(2,38) = 8.652$, $p < 0.01$. Emotion recognition accuracy for all conditions is presented in Fig. 2.

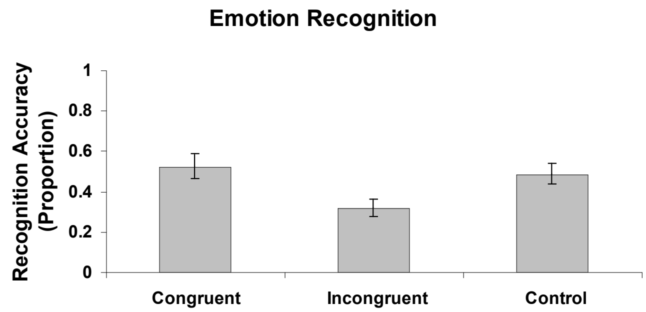

Figure 2. Emotion recognition accuracy (error bars indicate standard error).

A post hoc analysis of paired-samples *t* tests was conducted to further explore the patterns between conditions. As expected, the post hoc analysis indicated that incongruent context-emotion video pairings (M = 0.317, SD = 0.185) resulted in significantly lower emotion recognition accuracy compared to the congruent context-emotion pairs $t(1,19) = -3.526$, $p < 0.01$ and the control condition (emotion videos only) $t(1,19) = -3.651$, $p < 0.01$. For example, participants were able to interpret the robot's "happy" expression more accurately when displayed in the task context of finding dirt on the ground, as opposed to an incongruent display when unsuccessful docking or becoming stuck between two objects. No significant finding was found comparing the control (emotion videos only) (M = 0.485, SD = 0.233) and the congruent context-emotion pairings (M = 0.523, SD = 0.277).

B. Context Recognition

To determine if context recognition accuracy differed as a function of context-emotion congruency, a one-way repeated measures ANOVA was conducted. The ANOVA revealed that the context recognition mean scores differed significantly, with $F(2,38) = 15.32$, $p < 0.001$. Context recognition accuracy for all conditions is presented in Fig. 3.

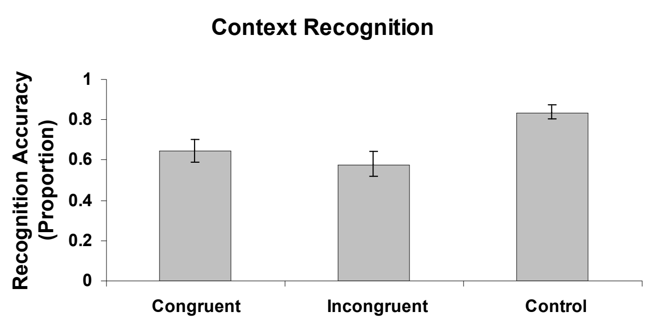

Figure 3. Context recognition accuracy (error bars indicate standard error).

Paired-samples *t* tests were conducted post hoc to further explore the patterns between conditions. The post hoc analysis indicated that control condition of the task context video (M = 0.836, SD = 0.164) resulted in significantly higher emotion recognition accuracy when compared to both the congruent context-emotion pairs $t(1,19) = 4.497$, $p < 0.001$ and the incongruent context-emotion pairs $t(1,19) = 4.396$, $p < 0.001$. In contrast to our hypothesis, the comparison between the congruent (M = 0.643, SD = 0.257) and incongruent context-emotion pairings (M = 0.578, SD = 0.285) did not yield a significant result. For example, participants were able to interpret the task context "sensing dirt on the ground" more accurately when that context was viewed in isolation, as opposed to paired with either an congruent ("happy") or incongruent ("sad") emotion.

V. DISCUSSION

This study sought to investigate the relationship between task context and emotion, as it specifically relates to HRI. We investigated pairs of task context and emotion videos, some pairs congruently matching users' expectations, while others did not. As a control measure, participants also viewed task context and emotion videos individually (not paired). By measuring both emotion and context recognition, the results, in part, supported our hypothesis that incongruent context-emotion pairings do influence recognition accuracy.

Particularly regarding emotion recognition the results support our hypothesis [H1]. Incongruent context-emotion pairings (e.g., instances where the robot depicted an emotional response to a context dissimilar to what most individuals would expect) influenced emotion recognition, resulting in lowered recognition accuracy.

Context recognition, however, only partially supported our hypothesis [H2]. The incongruent context-emotion pairings did results in lower context recognition accuracy when compared to the control (e.g., context videos only). Conversely, the incongruent video pairings did *not* differ significantly from the congruent pairings. This finding may suggest that the addition of emotion, either congruent or incongruent, to a task context may actually influence the participant to 'second guess' their identification of the context. While this finding is not what we expected, we would like to stress that further investigation is needed to fully understand the nature of emotion's role in context identification, as it relates to HRI.

In summary, the results suggest a directional relationship between context and emotion, but the specifics of that relationship may not appear to be necessarily as we expected. Nevertheless, our findings are important for designers of emotional robots, for they illustrate that the appropriateness of a robot's emotional response to a task context is critical in the user correctly recognizing the intended emotional message the robot is trying to convey.

VI. CONCLUSION AND FUTURE WORK

Our efforts here focused upon a specific robot, performing certain actions in response to a limited variety of situations, judged for efficacy by a restricted set of observers. The goal of our experiment was to investigate the interplay of context and emotion for a non-anthropomorphic robot.

Concerning context, we note that the contextual observation–that a robot forms a part of its own context–requires that the robot somehow must be aware of its environment and itself in relationship with that environment when selecting an appropriate behavior or trying to elicit an appropriate response. Additionally, and apart from the robot itself, the human user or observer of a robot may develop or already possess a set of expectations of behavior of the robot through either direct experience or indirect evidence. These expectations would frame the lens through which the context of the robotic interaction occurs. In this sense, the observer is a part of the context as well.

We shortly will undertake a course of study to take a more careful consideration the ages and general experience level of observers. We anticipate that by conducting such cross-sectional studies we might construct a mechanistic accounting of the amplitude of emotional expression necessary for recognition.

An overarching goal of our social robotics research is to develop a general purpose repertoire for expression. As we noted above, in the present study, we limited our experiments deliberately to using the Roomba. While we noted that evocativeness in robots is likely idiosyncratic to each robot, we feel it important to investigate those nuances afforded by intrinsic robot design by repeating the present experiment using a variety of robot body shapes and sizes. Simultaneously, the use of differing robots in an analogous study would afford both exploration of a broader variety of gestures, expressions and behaviors, and observation of such robots in a wider variety of anticipated commonplace occurrences.

The research we report in this paper is among the initial steps we are taking toward the development of an emotional repertoire for robots. We are presently at work to incorporate these findings into a general emotional intelligence system, which we expect will be aware of its context and will respond in an appropriate social manner under a variety of conditions.

ACKNOWLEDGMENT

The authors wish to express their gratitude toward Dr. Ashok Goel, Dr. Arthur D. Fisk, and Dr. Wendy A. Rogers. Also, we thank the students of the Socially Intelligent Machines Lab, the Design Intelligence Lab, and Intelligent Narrative Computing Lab for their many helpful comments and support.

REFERENCES

[1] C. Nass, B. Frogg, and Y. Moon, "Can computers be teammates?," International Journal of Human-Computer Studies, vol. 45, Dec. 1996, pp. 669-678.

[2] C. Nass and Y. Moon, "Machines and Mindlessness: Social Responses to Computers," Journal of Social Issues, vol. 56, Spring. 2000, pp. 81-103.

[3] C. Nass, Y. Moon, B.J. Fogg, and B. Reeves, "Can computer personalities be human personalities?," International Journal of Human-Computer Studies, vol. 43, 1995, pp. 223-239.

[4] J. Sung, L. Guo, R.E. Grinter, and H.I. Christensen, ""My Roomba is Rambo": Intimate Home Appliances," UbiComp 2007: 9th International Conference on Ubiquitous Computing, Innsbruck, Austria: 2007.

[5] J. Forlizzi, "How robotic products become social products: an ethnographic study of cleaning in the home," ACM/IEEE International Conference on Human-Robot Interaction, Arlington, Virginia, USA: 2007, pp. 129-136.

[6] J. Sung, R.E. Grinter, H.I. Christensen, and L. Guo, "Housewives or technophiles?: understanding domestic robot owners," ACM/IEEE International Conference on Human-Robot Interaction, Amsterdam, The Netherlands: 2008, pp. 129-136.

[7] J. Sung, R.E. Grinter, and H.I. Christensen, “"Pimp My Roomba": designing for personalization,” Conference on Human Factors in Computing Systems, Boston, MA, USA: 2009, pp. 193-196.

[8] J. Sung, H.I. Christensen, and R.E. Grinter, “Robots in the wild: understanding long-term use,” ACM/IEEE International Conference on Human-Robot Interaction, La Jolla, California, USA: 2009, pp. 45-52.

[9] G. Hoffman and C. Breazeal, “Collaboration in human-robot teams,” AIAA 1st Intelligent Systems Technical Conference, Chicago, IL, USA: 2004.

[10] World Robotics: Statistics, Market Analysis, Case Studies and Profitability of Robot Investment, 2005, United Nations, Geneva, Switzerland: United Nations Publications, 2005.

[11] M. Siemer, I. Mauss, and J.J. Gross, “Same situation-different emotions: How appraisals shape our emotions,” Emotion, vol. 7, 2007, pp. 592-600.

[12] C.L. Breazeal, Designing Sociable Robots, The MIT Press, 2004.

[13] P. Ekman and W.V. Friesen, Unmasking the face, Englewood Cliffs, NJ, USA: Prentice-Hall, 1975.