The Interplay of Context and Emotion for Non-Anthropomorphic Robots

Bryan Wiltgen, Jenay M. Beer, Keith McGreggor, Karl Jiang, and Andrea Thomaz

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 context and emotion recognition of a non-anthropomorphic robot, the iRobot Roomba. Considerations, implications, and future work are discussed.

I. INTRODUCTION

Household robots are becoming more commonplace in the world [1]. 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 [2]-[4]. 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 [5]-[7];
- Users develop emotional connections to their robot, saying that they "love it" or considering it "part of the family" [8]; 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 [9].

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 [10]. 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.

Our long-term research agenda is to improve the emotional intelligence and emotional evocativeness of a robot by constructing an emotional repertoire for it, 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, which we define as situations where the Roomba is interacting with its environment or requires human intervention (c.f., Table 1), could be improved by affording to the Roomba an extensive emotional repertoire. We believe this repertoire can be achieved programmatically by way of deliberately chosen actions rather than through unintentional or ambiguous assignment based on the robot’s intrinsic design.

In this work we investigated the significance context has upon a person’s ability to accurately interpret a robot’s selected emotional expression. In particular, we intend to illustrate that an emotional display is most accurately interpreted when expressed in an appropriate context, as compared to an inappropriate context or without context.

Manuscript received July 18, 2010.
Bryan Wiltgen is with the School of Interactive Computing at the Georgia Institute of Technology, Atlanta, GA 303302 USA. E-mail: bryan.wiltgen@gatech.edu. Phone: 404-218-8763.
Jenay M. Beer is with the School of Psychology at the Georgia Institute of Technology, Atlanta, GA 30332 USA. E-mail: jbeer@gatech.edu.
Keith McGreggor is with the School of Interactive Computing at the Georgia Institute of Technology, Atlanta, GA 30332 USA. E-mail: keith.mcgreggor@gatech.edu.
Karl Jiang is with the School of Interactive Computing at the Georgia Institute of Technology, Atlanta, GA 30332 USA. E-mail: kjiang@gatech.edu.
Andrea Thomaz, PhD. is with the School of Interactive Computing at the Georgia Institute of Technology, Atlanta, GA 30332 USA. E-mail: athomaz@cc.gatech.edu.
altogether. This has important implications for the design and evaluation of emotional robots, particularly non-anthropomorphic emotional robots.

II. THE INTERPLAY OF CONTEXT AND EMOTIONS

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, a designer may inadvertently develop emotive behaviors that may have a specific interpretation in one context, but are interpreted entirely differently within another. In addition, emotive behaviors that are inappropriately matched with a context might cause users to interpret that particular context incorrectly.

In everyday social encounters, people make determinants about each others emotional stance based upon a variety of social cues, such as emotion and situational context. In the literature, it is clear that there is a relationship between human expressed emotion and context; however, contrary theoretical predictions provide opposing views as to the exact nature of this interaction. One perspective suggests facial emotional dominance; that is, emotion information displayed by human facial expression overrides any expectations derived from the situational context [11]. However, more recent studies suggest that judgments of emotion are influenced by the contextual situation [12]-[14] particularly increasing speed of categorization if the emotion target and context are congruent.

We seek to know whether (a) emotion recognition is influenced by contextual information or (b) context recognition influenced by emotional information? What is the nature of the interplay between contextual and emotional information? It may even be that (c) the two aspects are mutually dependent. While it is generally understood that context and emotion are related [15], 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 appearance. 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 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 should be designed with a notion of how the robot’s present context might cause those selected behaviors to be interpreted emotionally.

B. Research Questions

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

· How can the context within which a person observes a robot influence emotion recognition?

· 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 inappropriate with respect to the users’ expectations, which variable 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 hypothesize that the interplay between emotion and context is such that inappropriate pairings of emotion-context will result in less accurate emotion recognition, and similarly, that inappropriate emotion-context pairings will also 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 younger 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. All participants indicated some experience with computing and/or robotics. Additionally, all participants indicated having some experience with the roomba specifically (i.e., they have, at minimum, previously seen the robot in person). The participants volunteered, and were not compensated for their time.

We acknowledge that our sample has a strong male bias and thus any claims we make herein require further experimentation to verify generality across genders. However, our sample was randomly selected from among the Georgia Institute of Technology population.

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 using 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.

![Secured experimental website](image)

**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). Contexts were identified as situations where either the Roomba was interacting with the environment (e.g., sensing dirt on the ground) 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. Videos were used instead of direct interaction with the robot to provide experimental control and repeatability across subjects.

A pilot test was conducted where participants were asked to empirically label the content of the videos. The pilot consisted of 10 participants who each viewed a series of emotion videos and, separately, a series of context videos.

Participants saw each video 3 times. The most highly agreed upon emotive behaviors and contexts were selected for our main experiment. For the contexts we selected, an average of 87.5% trials (across participants) were consistently labeled. For the emotive behaviors we selected, an average of 73.3% trials (across participants) were consistently labeled. 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 questionnaire was administered which asked the same pilot participants to identify emotion(s) that they would expect the Roomba to demonstrate in response to possible contexts. We refer to these as “appropriate” emotions in relation to a context. The most highly agreed upon selections (with 60+% agreement) were used to form the appropriate emotion-context pairings (see Table 2).

![Description of Emotion and Context Videos](image)

**Table 1**

<table>
<thead>
<tr>
<th>Context Videos</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bump</td>
</tr>
<tr>
<td>Dirt</td>
</tr>
<tr>
<td>Dock (unsuccessful)</td>
</tr>
<tr>
<td>Edge</td>
</tr>
<tr>
<td>Lift</td>
</tr>
<tr>
<td>Stuck</td>
</tr>
</tbody>
</table>

**Emotion Videos**

<table>
<thead>
<tr>
<th>Emotion</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Anger</td>
<td>Moving forward in a staccato manner while blinking a red light</td>
</tr>
<tr>
<td>Fear</td>
<td>Moving backward at a slow speed</td>
</tr>
<tr>
<td>Happiness</td>
<td>Display a constant blue light while producing a chirping, uptempo sound</td>
</tr>
<tr>
<td>Sadness</td>
<td>Moving forward at a slow speed while producing a low vibrato sound</td>
</tr>
<tr>
<td>Surprise</td>
<td>Starting forward, then moving backward at a fast speed while producing a rapid high-pitched beeping sound</td>
</tr>
</tbody>
</table>

![Description of Context-Emotion Pairings](image)

**Table 2**

<table>
<thead>
<tr>
<th>Appropriate Pairs</th>
<th>Inappropriate Pairs</th>
</tr>
</thead>
<tbody>
<tr>
<td>Context</td>
<td>Emotion</td>
</tr>
<tr>
<td>Bump</td>
<td>Surprise</td>
</tr>
<tr>
<td>Dirt</td>
<td>Happy</td>
</tr>
<tr>
<td>Dock (unsuccessful)</td>
<td>Anger</td>
</tr>
<tr>
<td>Edge</td>
<td>Fear</td>
</tr>
<tr>
<td>Lift</td>
<td>Surprise</td>
</tr>
<tr>
<td>Stuck</td>
<td>Anger</td>
</tr>
</tbody>
</table>
The experimenters then determined the “inappropriate” pairs by selecting an emotion opposite of the appropriate pair. Using Breazeal’s notion of arousal, stance, and valence space for emotions [16], we selected the inappropriate emotion by choosing an emotion on the opposite side of the space from the appropriate emotion. For example, if the appropriate emotion were happy (slightly high arousal, positive valence, neutral stance), we would select sad (low arousal, negative valence, and neutral stance) as the inappropriate emotion. In cases where this method was ambiguous, we leveraged our pilot data to guide our selection.

D. Design

A within-subjects design was conducted. Each participant viewed Appropriate Context-Emotion pairs, Inappropriate 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, they were prompted 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 6 Basic Emotions (anger, disgust, fear, happiness, sadness, and surprise) [17] 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:

· Appropriate and inappropriate pairs were presented and participants were asked to identify the emotion the Roomba demonstrated, if any;

· Appropriate and inappropriate 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 across participants using a partial Latin-Square to reduce carry-over effects. Each participant viewed 6 blocks, 2 of each type.

F. Hypotheses

Through careful pilot testing, the experimenters created videos of contexts and emotions that were readily identified by participants. The experimenters then systematically created three conditions: appropriate context-emotion pairings, inappropriate context-emotion pairings, and a control (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 appropriate video pairings and the control, for either emotion recognition or context recognition. However, the inappropriate context-emotion pairings were expected to influence recognition accuracy, which generated the two following hypotheses:

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

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

IV. RESULTS

Separate repeated measures analyses of variance (ANOVAs) were conducted to test the effects of emotion-context appropriateness 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 appropriateness. A statistically significant main effect was found, indicating that the emotion recognition accuracy mean scores differed by condition $F(2,38) = 8.652, p < 0.01$. Emotion recognition accuracy for all conditions is presented in Fig. 2.

![Emotion Recognition](image)

Fig. 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 inappropriate context-emotion video pairings ($M = 0.317$, $SD = 0.185$) resulted in significantly lower emotion recognition accuracy compared to the appropriate 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 context of finding dirt on the ground, as opposed to an inappropriate display when unsuccessful docking or becoming stuck between two objects. In fact, for the inappropriate context-emotion pairings, participants often mislabeled the emotion as neutral (i.e., no emotion present). No significant finding was found comparing the control (emotion videos only) ($M = 0.485$, $SD = 0.233$) and the appropriate context-emotion pairings ($M = 0.523$, $SD = 0.277$).

**Context Recognition**

To determine if context recognition accuracy differed as a function of context-emotion appropriateness, 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.

![Fig. 3. Context recognition accuracy (error bars indicate standard error).](image)

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 context video ($M = 0.836$, $SD = 0.164$) resulted in significantly higher context recognition accuracy when compared to both the appropriate context-emotion pairs ($t(1,19) = 4.497$, $p < 0.001$) and the inappropriate context-emotion pairs ($t(1,19) = 4.396$, $p < 0.001$). In contrast to our hypothesis, the comparison between the appropriate ($M = 0.643$, $SD = 0.257$) and inappropriate context-emotion pairings ($M = 0.578$, $SD = 0.285$) did not yield a significant result. For example, participants were able to interpret the context “sensing dirt on the ground” more accurately when that context was viewed in isolation, as opposed to paired with either an appropriate (“happy”) or inappropriate (“sad”) emotion. When participants mislabeled a context, they often labeled the context as “don’t know”, suggesting that the pairing of context and emotion allowed for disambiguation.

V. **DISCUSSION**

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

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

Context recognition, however, only partially supported our hypothesis [H2]. The inappropriate context-emotion pairings resulted in lower context recognition accuracy when compared to the control (e.g., context videos only). Conversely, the inappropriate video pairings did not differ significantly from the appropriate pairings. This finding may suggest that the addition of emotion, either appropriate or inappropriate, to a 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 relationship between context and emotion, but the specifics of that relationship may not appear to be necessarily as we expected. Moreover, our findings illuminate how the relationship between emotion and context recognition seems to be more complex than hypothesized.

Nevertheless, these findings are important for designers of emotional robots. People are willing to apply social characteristics to robotic systems and social cues may improve human robot interaction and facilitate natural social interaction. However, our findings suggest that such social cues should be chosen carefully. These data illustrate that the appropriateness of a robot’s emotional response to a context is critical in the user correctly recognizing the intended emotional message the robot is trying to convey.

VI. **CONCLUSION AND FUTURE WORK**

The goal of our experiment was to investigate the interplay of context and emotion for a non-anthropomorphic robot, the iRobot Roomba.

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 emotional behavior or trying
to elicit an appropriate response. Additionally, the human user or observer of a robot may develop or already possess a set of expectations of behavior of the robot as a result of direct experience. These expectations would frame the lens through which the context of the robotic interaction may be perceived by the human. In this sense, the observer is a part of the context as well.

One overarching and ongoing goal of our social robotics research is to develop an emotional 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.

ACKNOWLEDGMENTS

The authors wish to 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