

Interactions with a Moody Robot

Rachel Gockley
rachelg@cs.cmu.edu

Jodi Forlizzi
forlizzi@cs.cmu.edu

Reid Simmons
reids@cs.cmu.edu

Carnegie Mellon University
5000 Forbes Avenue
Pittsburgh, PA 15213

ABSTRACT

This paper reports on the results of a long-term experiment in which a social robot's facial expressions were changed to reflect different moods. While the facial changes in each condition were not extremely different, they still altered how people interacted with the robot. On days when many visitors were present, average interactions with the robot were longer when the robot displayed either a "happy" or a "sad" expression instead of a neutral face, but the opposite was true for low-visitor days. The implications of these findings for human-robot social interaction are discussed.

Categories and Subject Descriptors

J.4 [Social and Behavioral Sciences]: psychology; I.2.9 [Artificial Intelligence]: robotics; H.1.2 [Models and Principles]: user/machine systems—*human factors*

General Terms

Human Factors, Design, Experimentation

Keywords

Human-robot interaction, social robots, psychology, emotions, moods, affective modeling

1. INTRODUCTION

Why might we want a robot to display emotions? For robots that interact socially with people, we believe that emotions may be key for allowing such interactions to occur smoothly and naturally. This paper presents a long-term study of how a robot's mood influences its social interactions with people. Over the nine-week study, we found that simply changing the robot's facial expression caused people to interact more or less with it, depending on which mood the robot displayed and on how familiar the person was with the robot.

Emotions play a major part in human interaction, and much research has been done to indicate the social nature

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

HRI'06 March 2–4, 2006, Salt Lake City, Utah USA.
Copyright 2006 ACM 1-59593-294-1/06/0003 ...\$5.00.

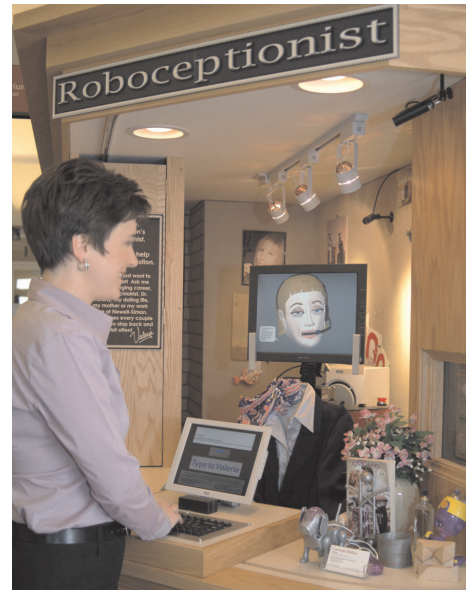


Figure 1: A visitor interacting with Valerie the Roboceptionist.

of emotions. Quite often, emotional reactions are caused by social interactions, influenced by societal and cultural norms, or used to communicate desires to other people [12]. What mood a person is in has a strong impact on how that person interacts with others [7]. In addition, people who are interacting may "catch" each other's emotions, taking on the other's emotional state [19].

A well-studied effect of human-computer interaction is that people tend to react to computers in the same way that they react to other people [14]. Furthermore, research indicates that this tendency carries over into long-term "relationships" with computers, such as students preferring particular computer terminals in a lab [18]. We believe that this tendency to form social relationships with computers also applies to robots, perhaps even more so. If that is the case, then people should respond to a robot's emotions as though the robot were human. However, very little research has been done to study this effect.

In recent years, the robotics community has seen a gradual increase in social robots, that is, robots that exist primarily to interact with people. Museum tour-guide robots [11, 3] and robots that interact with the elderly [9] demonstrate

not only the benefits of having robots interact with people, but also the need for the interactions to be smooth and natural. Many robots in the previously mentioned fields have incorporated at least some rudimentary emotional behaviors and have demonstrated their usefulness, particularly in terms of how willing people were to interact with the robots. Furthermore, robots such as Kismet [1], which is capable of infant-like interactions with humans, have been used to demonstrate the ability of people to interpret and react appropriately to a robot’s displays of emotions. None of these robots, however, have been used to investigate people’s reactions to the robot’s emotions in spontaneous, unattended interactions—in truly human-like situations.

The exploration of emotions in robots currently lags behind similar research in software agents. For example, both the Oz Project [15] and the Affective Reasoner [6] are implementations of virtual worlds populated by software agents, wherein the agents can detect and react to each others’ emotions. In neither of these cases, however, do the agents interact directly with humans. Embodied conversational agents often display emotions in their interactions with people [4], but the specific role of affect has yet to be addressed. The most developed robotic emotional model that we are aware of is the TAME architecture [10], which considers the four affective categories of personality traits, attitudes, moods, and emotions; however, this model has not yet been fully implemented or tested.

Our research team has previously worked on social robots, including Vikia [2] and Grace [17]. Currently, our platform for social robotics research is Valerie the Roboceptionist [8]. The mechanical base of the robot is an RWI B21r with an LCD “head” mounted on a pan-tilt unit, which allows the entire monitor to rotate. The robot is housed in a custom-built booth, near a high-traffic entrance to the School of Computer Science at Carnegie Mellon University. A keyboard and small monitor on the booth’s desk allow for human input with visual feedback. The LCD head displays a highly expressive, graphical face. Valerie can communicate by using speech synthesis with automatic lip syncing of the face. Because the face is so expressive, it can display a wide range of easily recognizable emotional expressions. In a collaboration with the Drama Department, Valerie has a personal history, including a complex, evolving storyline. Visitors to her booth can ask her questions about her life in order to hear the continuing story. For example, the major plot arc of the 2004-2005 academic year was that Valerie fell in love with a jukebox named Cal, but they were unable to marry legally due to their not having human status. These stories are intended to provide a reason for people to interact with Valerie even when they do not need her assistance in obtaining directions around campus. A picture of someone interacting with the robot is shown in Fig. 1.

2. HYPOTHESES

In what ways might a robot’s emotional expressions influence people’s perceptions and behaviors? We designed a study to examine the effects of relatively static mood expressions on the interactions that Valerie has with people. We hypothesized that even a relatively minor change in the robot’s facial expression would have a significant effect on how people interact with the robot; specifically:



Figure 2: The three mood expressions used in this experiment: neutral, positive, and negative (left to right). Differences include eyebrow and eyelid positions, mouth shape, and head angle.

1. We predicted that people would react to Valerie’s expressions in the same way that they do with humans. Specifically, we hypothesized that more people would interact, and would interact for longer periods, with the happy robot than with the neutral robot. Additionally, because people typically prefer not to interact with people who are depressed [16], we believed that fewer people would interact, and for shorter periods, with the sad robot than with the neutral robot.
2. We predicted that people would subjectively rate their enjoyment of Valerie in the same manner, enjoying interactions with the positive robot the most and the negative robot the least.
3. Occasionally, events at the University bring an unusually large number of visitors near Valerie’s booth for a week at a time. These weeks would likely have a higher than usual proportion of people who have no prior familiarity interacting with Valerie. During these “high” weeks, we predicted that interactions with the robot would exhibit little difference between the three conditions, as the novelty of the robot would outweigh the effect of the affective expressions.

3. PROCEDURE

The study was run over a total of nine weeks (non-contiguous), with the robot typically operating eight hours per day, five days per week. During each day, the robot displayed either her usual “neutral” mood or a positive or negative mood. We altered the moods across days and weeks both to match the robot’s life stories and to account for daily differences in interactions. The different moods were indicated by a change in facial expression, as shown in Fig. 2, as well as minor behaviors the robot performed during conversations while waiting for the visitor to type a statement. These behaviors included:

- In the neutral condition, Valerie either smiled and performed a single head nod or briefly glanced away from the visitor. These were the same behaviors that Valerie typically performed prior to this study.
- In the positive (happy) condition, Valerie either smiled and nodded as with neutral (but with a wider smile), or bounced her head from side to side in a seemingly happy, energetic motion.

- In the negative (sad) condition, Valerie either looked away from the visitor or appeared to sigh.

We pre-tested the facial expressions and behaviors and found that people could accurately determine the intended moods. To avoid bias due to incongruent mood displays, the positive mood was run during weeks in which the storyline was also positive, and the negative mood during weeks in which the storyline was negative. No other changes were made to the interaction structure. As described elsewhere [8], people can interact with Valerie by typing on a keyboard, and Valerie responds using automatic speech generation. Visitors have the option of swiping an ID card in Valerie’s cardreader to identify themselves to the robot, but the robot otherwise has no means of identifying interactors even if they had previously interacted with the robot.

During each day of the experiment, we used the laser rangefinder in the booth to detect and record the number of people who interacted with the robot. For each interactor, the robot logged the following: the time the interaction began, the number of seconds from the start of interaction until the visitor left the laser’s detection range (about two meters away from the keyboard), the number of lines of input typed to the robot during the interaction, and whether any other visitors interacted with the robot during the same time, such as when a group of people took turns interacting. To identify repeat interactors, people who swiped an ID card were matched against a database of people who had previously interacted with the robot.

At various points during the study, we approached visitors who had just interacted with the robot to conduct a brief anonymous verbal survey. Initially, we asked participants to rate the robot’s valence, arousal, and mood appropriateness, as well as their entertainment, on a 5-point scale. As the study progressed, we determined a need for additional data in the positive and neutral conditions, resulting in a second survey, which asked about the robot’s valence and arousal, how natural, likable, and entertaining the robot was, the clarity (or enigma) of the robot’s thoughts, and the person’s comfort and mood level with the robot, using a 7-point scale.

4. RESULTS

We considered two measures of how much people interacted with the robot: how many **seconds (S)** each person spent with the robot, and how many **lines of input (L)** each person typed. These two measures were highly correlated ($r = .73$) but had different enough meanings that we chose to analyze them separately. We considered the following independent variables in our analyses:

- **Affect (A)** refers to the affective condition of the robot: positive (happy), negative (sad), or neutral.
- Based on prior experience with Valerie, we had reason to suspect that the **time of day (T)** would affect people’s interactions with the robot. We grouped times into three categories: morning (before noon), afternoon (noon to 3pm), and evening (after 3pm).
- We also considered whether a person interacted with Valerie as part of a **group (G)** of interactors or individually. We classified a person as part of a group only if at least one other person also typed to the robot

Table 1: Days and number of interactors (I) in each condition, as well as average interactors per day (I/D), by high and low visitor traffic weeks.

<i>Condition</i>	Low Traffic			High Traffic		
	Days	I	I/D	Days	I	I/D
Neutral	9	458	50.9	7	483	69.0
Positive	7	410	58.6	8	636	79.5
Negative	8	380	47.5	4	312	78.0
Total	24	1248	52.0	19	1431	75.3

while that person was present. This distinction is due to the fact that the robot could not determine whether nearby people were part of a coherent group with the interactor if the others did not type anything.

Because we observed different interaction patterns between weeks of different visitor traffic levels, we conducted separate analyses for each level. For each of seconds and lines, we began with a one-way analysis of variance (ANOVA), modeling the main effects and all possible interaction effects of affect, time, and group. Given the large number of observations, we chose to remove higher-ordered interactions with F values less than 1 from the analysis. This allowed us in many cases to report on effects of single variables (i.e. main effects) without the added confusion of interactions between multiple variables. We will not discuss dropped terms.

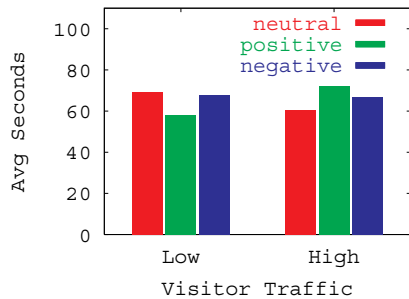
4.1 Low visitor traffic weeks

Few visitors came to the University during five of the nine weeks of the study. That is, these weeks had no major events that brought a large number of visitors to the University, so most of the people who interacted with the robot likely had some degree of familiarity with it. The robot was available for 24 days during these weeks, during which time a total of 1248 people interacted with it by typing at least one line of text. The number of interactors in each condition is given in Table 1.

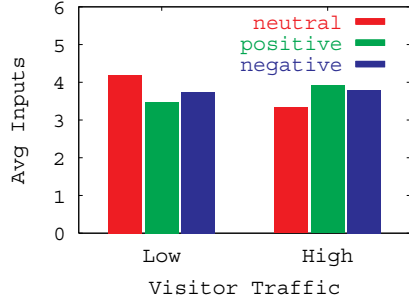
4.1.1 Seconds

After dropping terms as described above, our final model of seconds included the main effects of affect, time of day, and group, and the interaction of affect and time of day. This model predicted approximately 1.8% of the variance of seconds spent talking with the robot. The interaction between affect and time of day was not significant ($F[4, 1230] = 1.32$, n.s.). We found significant main effects for each of affect ($F[2, 1230] = 2.95$, $p = 0.05$), time of day ($F[2, 1230] = 2.92$, $p = 0.05$), and group ($F[1, 1230] = 8.55$, $p = 0.004$).

These main effects can be explained as follows. People interacted with Valerie for similar times in the neutral and negative conditions (neutral $M = 69.0$, negative $M = 67.7$), but for less time in the positive condition ($M = 57.9$), as can be seen in Fig. 3. Comparisons of mean values (contrast tests) showed that the neutral condition did not differ significantly from both affective conditions ($F[1, 1238] = 0.73$, n.s.). However, people interacted for significantly more time in the positive condition than in the neutral condition ($F[1, 1238] = 3.70$, $p = 0.05$). People in groups interacted for a significantly longer time than individuals (in-group $M = 70.3$, alone $M = 59.3$), as shown in Fig. 4. Time spent interacting varied significantly by time of day, with the longest interactions occurring in the morning ($M = 70.9$),



(a) Seconds spent with the robot



(b) Lines typed to the robot

Figure 3: Average seconds spent with and lines of input typed to the robot in each mood condition, compared across low and high visitor traffic weeks.

shortest in the afternoon ($M = 61.4$), and evening interactions lasting in between ($M = 66.6$).

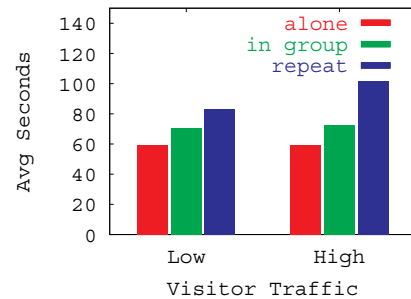
4.1.2 Inputs

We performed a similar analysis for lines, yielding a one-way ANOVA modeling the main effects of affect, time of day, and group, and the interaction effects of group and time and of affect and time. This model explained approximately 1.5% of the variance in inputs. The interaction of group with time of day was not significant ($F[2, 1236] = 1.06$, n.s.), nor was the interaction of affect with time of day ($F[4, 1236] = 1.21$, n.s.). Unlike seconds, there were no significant main effects of group ($F[1, 1236] = 2.43$, n.s.) or of time of day ($F[2, 1236] = 2.20$, n.s.). Affect continued to have a significant main effect ($F[2, 1236] = 3.10$, $p = 0.05$).

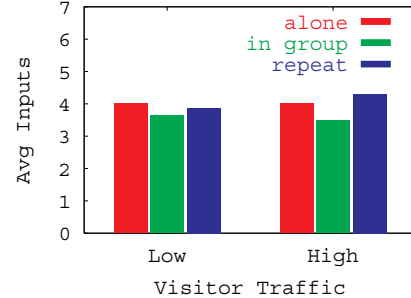
Contrast tests showed that people typed significantly more lines to Valerie in the neutral condition than in either emotional condition (neutral $M = 4.19$, positive $M = 3.49$, negative $M = 3.74$; $F[1, 1236] = 4.59$, $p = 0.03$). The number of lines typed did not differ between the positive and negative conditions ($F[1, 1236] = 1.49$, n.s.).

4.2 High visitor traffic weeks

We classified the remaining four weeks as “high visitor traffic” weeks. During these weeks, unusually high numbers of University visitors passed through the building in which the robot is located, resulting from events such as summer classes for visiting high school students and scheduled visits for incoming undergraduates and their families. The robot was available for 19 days during these weeks, during which time 1431 people interacted with it (Table 1).



(a) Seconds spent with the robot



(b) Lines typed to the robot

Figure 4: Average seconds spent with and lines of input typed to the robot by interactors in a group, interactors alone, and repeat interactors, compared across low and high visitor traffic weeks.

Significantly more people interacted with the robot per day during the “high visitor traffic” weeks than during the “low” weeks ($F[1, 41] = 19.44$, $p < .001$), which is consistent with the theory that many people who interacted with the robot during these weeks were visitors who were likely unfamiliar with the robot.

4.2.1 Seconds

As with low-traffic weeks, we performed a one-way ANOVA on seconds, modeling all main and interaction effects of affect, time of day, and group. The three-way interaction effect did not satisfy the $F < 1$ heuristic for removal discussed earlier, so the model was not changed. This model explained approximately 2.8% of the variance in seconds. The three-way interaction was not significant ($F[4, 1413] = 1.12$, n.s.), nor were the two-way interactions between affect and time of day ($F[4, 1413] = 1.18$, n.s.) or between affect and group ($F[2, 1413] = 0.81$, n.s.). Unlike the “low” weeks, there was a significant interaction effect between group and time of day ($F[2, 1413] = 3.23$, $p = 0.04$). Additionally, there were significant main effects of affect ($F[2, 1413] = 4.18$, $p = 0.02$), time of day ($F[2, 1413] = 3.35$, $p = 0.04$), and group ($F[1, 1413] = 15.0$, $p < 0.001$).

We performed contrast tests to describe the effects of affect. In contrast to the “low” weeks, people in the neutral condition interacted for a significantly shorter period of time than in either emotional condition (neutral $M = 60.1$, positive $M = 72.4$, negative $M = 66.7$; $F[1, 1413] = 4.53$, $p = 0.03$). The number of seconds spent interacting in the

negative and positive conditions were not significantly different ($F[1, 1413] = 1.64$, n.s.).

Holding all other terms constant, people who were within a group of interactors spent a longer amount of time with the robot (in-group $M = 72.6$, alone $M = 59.0$), as with low-traffic weeks. Additionally, people overall tended to interact for shorter periods of time in the afternoon than in either the morning or evening (morning $M = 67.5$, afternoon $M = 62.7$, evening $M = 72.5$), though this relationship was not significant ($F[1, 1413] = 2.96$, n.s.). Groups did not follow the same interaction patterns throughout the day as did individuals; a contrast test indicated a downward linear trend for groups and an upward linear trend for individuals (interaction $F[1, 1413] = 4.61$, $p = 0.03$).

4.2.2 Inputs

After removing terms with $F < 1$ from the full factorial analysis, our final model was a one-way ANOVA modeling the main effects of affect, time of day, and group, along with the interactions between group and time of day, and between affect and time of day. This model explained approximately 1.9% of the variance in the number of inputs typed. There were no main effects of group ($F[1, 1413] = 1.61$, n.s.) or of affect ($F[2, 1413] = 2.56$, n.s.), but there was a main effect of time of day ($F[2, 1413] = 5.42$, $p = 0.005$). There were no significant interactions between affect and time of day ($F[4, 1413] = 0.37$, n.s.) or between group and time of day ($F[2, 1413] = 2.72$, n.s.).

While the main effect of affect was not significant, contrast tests indicated that people typed significantly fewer lines to the robot in the neutral condition than in either emotional condition (neutral $M = 3.33$, positive $M = 3.92$, negative $M = 3.79$; $F[1, 1419] = 3.81$, $p = 0.05$). The positive and negative conditions did not differ ($F[1, 1419] = 0.37$, n.s.).

A contrast test showed that people tend to type more lines to the robot as the day progresses ($M = 3.38$ in the morning, $M = 3.41$ in the afternoon, and $M = 4.19$ in the evening; $F[1, 1419] = 6.23$, $p = 0.01$). While group did not cause a significant main effect, there was a slight interaction effect between group and time of day. The interaction between G and T on L followed the same trend as described for S above (interaction $F[1, 1419] = 5.38$, $p = 0.02$).

4.3 Repeat visitors

Of particular interest are those people who have interacted with Valerie multiple times, thus forming a “relationship” with her. Without using face detection or otherwise collecting personal information, however, the robot has no way of determining whether any particular person has interacted with it previously unless that person actively chooses to swipe an ID card on each visit. Due to hardware failures, the cardreader was not available every day that the robot was operating. Over all nine weeks, 76 of the interactors were identified in this manner as previous visitors to the robot. A higher number of visitors per day returned to the robot during the high visitor traffic weeks than during the low traffic weeks (see Table 2).

Because of the imbalance of repeat interactors across the conditions, we were unable to run a full factorial analysis as we had previously. Instead, we performed an ANOVA for the seconds spent interacting as well as the number of lines typed, modeled in terms of whether the interactor was a “repeat” and whether it was a high or low traffic week. No

Table 2: Days the cardreader was running (D), number of repeat interactors (R), as well as average repeat identifiers per day in each condition.

Condition	Low Traffic			High Traffic			Overall R/D
	D	R	R/D	D	R	R/D	
Neutral	8	6	0.75	7	24	3.43	1.88
Positive	7	6	0.86	8	26	3.25	2.00
Negative	5	3	0.60	4	11	2.74	1.27

Table 3: Average ratings of the robot’s valence and arousal from Survey 1. All variables were measured on a scale of 1–5. $N = 62$.

Question	Neutral		Positive		Negative	
	M	SD	M	SD	M	SD
Valence	2.73	1.03	3.00	1.15	2.50	0.86
Arousal	2.50	0.88	3.06	1.24	2.45	1.14

effects were significant, though repeat interactors tended to spend longer interacting with the robot ($M = 97.8$ seconds for repeat interactors, $M = 65.2$ overall). That is, repeat interactors spent more seconds with the robot than the average over all interactors, but otherwise did not differ.

Additionally, we analyzed the variance of both the number of seconds with the robot and the number of lines typed, modeling the effects of affect, time of day, and group. We did not consider interaction effects, due to the aforementioned imbalance. For seconds, all variables had $F < 1$, indicating that none had an effect on the time spent with the robot. For inputs, neither condition nor time of day was significant, but repeat interactors in groups typed significantly fewer lines to the robot than those not in groups ($M = 5.85$ for alone, $M = 3.56$ for groups; $F[1, 70] = 4.83$, $p = 0.03$).

4.4 Survey 1

The two surveys were designed to measure visitors’ subjective experiences of the robot. A total of 62 responses to the first survey were collected over two weeks: 16 on positive days, 22 on negative days, and 24 on neutral days. This survey measured participant’s ratings of how positive or negative the robot was (valence), the robot’s energy or excitation level (arousal), appropriateness of mood to what was said (mood congruency), and entertainment. Initially, we placed the surveys on the robot’s booth with a sign requesting interactors to complete the survey. As this collection method proved insufficient, the remainder of the surveys were collected by hand. The “self-survey” condition was included as an independent variable in the following analyses.

Valence and arousal were correlated ($r = 0.43$, $p < 0.001$), but no other measures were significantly correlated (at the $p < 0.01$ level), indicating that each question measured a different aspect of the visitor’s experience. When analyzed individually, none of the four questions (i.e. valence, arousal, mood congruency, and entertainment) differed significantly between the three mood conditions. However, by analyzing people’s ratings of the robot’s valence and arousal with a multivariate analysis of variance (MANOVA), we found that people were able to differentiate between the mood conditions ($F[2, 58] = 2.39$, $p = 0.10$). The positive condition had the highest valence/arousal ratings, negative had the lowest, and neutral fell in between, as shown in Table 3. The

Table 4: Results from Survey 2, by robot condition. Variables were measured on a scale of 1–7. $N = 61$.

Question	Neutral		Positive		F	p
	M	SD	M	SD		
Robot valence	3.91	1.06	4.27	1.43	1.08	0.30
Robot arousal	3.00	0.93	3.76	1.23	6.57	0.01
Natural	3.46	1.14	4.08	1.28	3.75	0.06
Likable	4.58	1.18	4.30	1.29	0.77	0.38
Entertainment	4.46	1.18	3.97	1.54	1.73	0.19
Enigmatic	5.08	1.21	4.22	1.57	5.29	0.03
Comfort	5.42	1.28	5.08	1.26	1.02	0.32
Person valence	5.33	1.24	4.49	1.33	6.25	0.02

differences were marginal, but this may be due to the small sample size or the granularity of the scales used.

Neither appropriateness nor entertainment differed significantly in the different mood conditions (appropriateness $F[2, 58] = 1.68$, n.s.; entertainment $F[2, 58] = 1.62$, n.s.). Overall, respondents responded near the scale’s midpoint for both the appropriateness of the robot’s mood ($M = 3.63$, $SD = 1.27$) and their entertainment ($M = 3.19$, $SD = 1.27$). The presence of the experimenter had an effect only on the rating of appropriateness, with individuals responding on their own rating the robot’s mood as less appropriate (self-survey $M = 3.00$, face-to-face $M = 3.78$; $F[2, 58] = 6.99$, $p = 0.01$).

4.5 Survey 2

We collected 61 responses to the second survey: 37 on positive days, and 24 on neutral days. The survey was designed to investigate the low interaction times in the positive condition during low traffic weeks (which was contrary to Hypothesis 1), and so no responses were collected on negative days. As discussed, this survey contained questions regarding both the robot’s and the respondent’s states. Results for each question are shown in Table 4.

The ratings of the robot’s valence, arousal, naturalness, and likability were pairwise correlated ($r > 0.3$, $p \leq 0.01$ for all pairs). The participant’s valence was correlated with the ratings of the participant’s comfort, the robot’s valence, likability, and entertainment ($r > 0.3$, $p < 0.01$ for each). Additionally, entertainment and likability were correlated ($r > 0.4$, $p < 0.001$). No other measures were significantly correlated.

As with the first survey, we computed a MANOVA on the robot’s valence and arousal. Respondents rated the robot’s valence/arousal as significantly higher in the positive condition than in the neutral condition ($F[1, 59] = 4.22$, $p = 0.04$). They found the positive robot significantly less enigmatic and slightly more natural than the neutral robot. There were no significant differences in ratings of the robot’s likability, entertainment, or participant’s comfort. Surprisingly, given the tendency mentioned earlier for people to “catch” others’ emotions, participant valence was significantly lower in the positive condition.

To further analyze the participant valence rating, we compared each participant’s valence to his or her rating of the robot’s valence (Fig. 5). A MANOVA showed a significant interaction effect ($F[1, 59] = 10.31$, $p = 0.002$). Visitors in the positive condition tended to rate the robot as just above the scale’s midpoint ($M = 4.27$, $SD = 1.43$), and rated

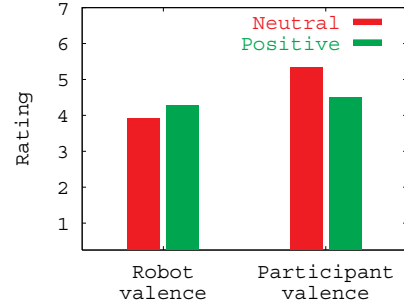


Figure 5: Average participant and robot mood ratings compared across conditions, from Survey 2.

themselves similarly ($M = 4.49$, $SD = 1.33$). Visitors in the neutral condition, however, tended to rate the robot’s mood as just below the scale’s midpoint ($M = 3.92$, $SD = 1.06$), and rated their own mood much higher ($M = 5.33$, $SD = 1.24$).

5. DISCUSSION

Despite our expectations, aspects of each hypothesis were disconfirmed. Though not always in the directions we expected, we did encounter multiple factors that related to how long people interact with Valerie and how much they say to her. These factors include: the affective condition of the robot, the time of day, whether a person interacted alone or as part of a group, whether the week included a large number of visitors to the University, and whether the person was identified as a repeat interactor.

Both surveys showed that people who interacted with the robot could distinguish between the robot’s moods, indicating that our manipulation was successful. Valence and arousal ratings were strongly correlated, as expected—happiness has high arousal and high valence, whereas sadness has low arousal and low valence. The robot’s mood did have an effect on people’s interactions with it, but the effect varied across high and low visitor traffic weeks. The differences were fairly large—on the order of ± 10 seconds, while interactions averaged only about a minute. (The differences between numbers of inputs followed the same trend.) As we expected, more people interacted with the positive robot, and fewer with the negative robot, during the low visitor traffic weeks. Contrary to Hypothesis 1 (that people would interact longest with the positive robot and shortest with the negative), however, people interacted for significantly less time with the positive robot during those weeks. We believe that this finding can be explained by relying on the theory of common ground [5]. Past studies have shown that people’s perception of common ground with robots results in shorter, more “efficient” speech [13]. Smiling carries a certain amount of conversational content, so people may not have felt that they needed to interact as long with the positive robot to come to the same level of understanding as when the robot did not smile. This theory is supported by the results of Survey 2; participants found the positive robot to be significantly less enigmatic—that is, easier to understand—than the neutral robot. Additionally, participants found the positive robot more natural than the neutral robot, which could also contribute to a sense of common ground.

During high-traffic weeks, in contrast, visitors interacted the least in the neutral condition and the most in the positive condition. Since the interactions did differ across conditions, even during the high visitor traffic weeks, Hypothesis 3 (that the mood would have little effect on high visitor traffic weeks) was disconfirmed. One possible explanation for this result may be that the moody robot appeared more interesting; perhaps any form of “emotion” in a robot may be compelling enough to sustain interactions.

We used both surveys to analyze Hypothesis 2 (that people would enjoy the positive robot the most and the negative robot the least). While we expected that people’s subjective rating of the robot’s entertainment would differ across the emotional conditions, this was not the case. Rather, participants in all conditions rated their interaction with the robot at a neutral level. This may be because the mood expressions were static; perhaps the mood needs to change more dynamically in order to affect people’s entertainment.

Though people’s entertainment did not change, Survey 2 showed an interesting interaction between how participants rated the robot’s mood and how they rated their own mood—people rated themselves as being in a considerably better mood when the robot was neutral. It may be the case that people rated their own mood in direct comparison to how they saw the robot. That is, people may have seen the positive robot as being relatively happy and then rated themselves similarly, but they may have considered themselves to be much happier than the neutral robot.

The typical behavior of people interacting alone versus people interacting in groups remained relatively consistent throughout the study. In particular, people who interacted in a group spent more time interacting, but typed slightly fewer lines, than people who interacted alone. This may imply that groups of people are likely to interact among themselves as well as with the robot, thus increasing their time with the robot but not necessarily increasing the number of exchanges they have with the robot. This finding corresponds to our general observations of people interacting with the robot.

Interestingly, the behavior of repeat interactors was not consistent across the weeks. While “repeaters” interacted for significantly longer than others, their interactions were longer during the high visitor traffic weeks than during the low weeks. Additionally, more repeat interactors visited the robot during the high visitor traffic weeks. One explanation may be that some of the University visitors during the “high” weeks were actually present multiple days, and may have swiped an ID card during their interactions with the robot. For example, the high school students taking classes in the building might have repeatedly interacted with Valerie. However, those students would still have had less experience with the robot than most University affiliates, and so they may have interacted differently than other repeaters.

Furthermore, the mood condition had no significant effect on the interaction times of repeat visitors. We consider this finding to be rather surprising, as we felt that people who interacted with the robot repeatedly would be more likely to notice and react to the different moods. One difference is notable, though: fewer repeat visitors interacted with the robot on negative days, and more on positive days, than on neutral days. This may imply that those people most familiar with the robot chose whether to interact with the robot at all based on her facial expression. Anecdotally, one

member of the University community expressed a concern for Valerie’s well-being during one of the negative periods, stating that she had been avoiding interacting with the robot as a result of Valerie’s expression!

Finally, the time of day during which an interaction occurred also related to differences in interaction patterns. Overall, people interacted for the least amount of time during the afternoon. While there are many possible explanations for this, the most likely is that people were busiest in the afternoon, and thus less likely to interact for long periods of time. An interaction effect between time of day and group was found only on high visitor traffic weeks: individuals tended to stay with the robot longer as the day progressed, whereas groups spent the most time in the mornings. It may be that individuals have more free time to interact with the robot later in the day, whereas groups—such as tour groups or families—might become rushed as the day progressed, leading to the decrease in interaction times (and inputs).

One obvious criticism of this analysis is that the amount of variance explained by our models is extremely small—on the order of only one or two percent. Presumably, then, many other factors contribute to how much people interact with the robot. Much of the variance can be explained by covarying time with inputs because the two measures are so highly correlated. This was not done as it would have added even more complexity to the analysis. Additionally, while the R^2 values of the models we used are admittedly low, we note that the differences found were on the order of 10 seconds with regard to minute-long interactions—a 15% increase or decrease. The low explanatory power of the overall model may just highlight the wide individual differences in how people interact with Valerie.

Another potential shortcoming of our analysis is that we do not currently have any means (beyond the cardreader) of identifying whether any particular person has interacted with the robot in the past, or even been familiar with the robot prior to his or her interaction during this study. That is, we do not have solid evidence that the differences between high and low visitor traffic weeks are primarily due to the visitors. However, we do know that the weeks we considered “high visitor traffic” did have an unusually high number of visitors to the University. Those same weeks also had more interactors per day than usual, as well as different interaction patterns than the other weeks. Thus, we feel that it is valid to assume that those weeks had a much larger proportion of people unfamiliar with the robot interacting with it, in comparison to the “low visitor traffic” weeks.

6. CONCLUSION

This research has many implications for the design of social robots. Our study showed that simply changing the apparent mood of a robot had a strong effect on people’s interactions with it. In particular, people who interacted with the robot during weeks of little visitor traffic at the University tended to interact for shorter periods with the positive robot, whereas people during high visitor traffic weeks interacted the most with the positive robot and least with the neutral. If we assume that the differences across weeks are primarily caused by people who were previously unfamiliar with the robot, then one interpretation of our results is that people who are familiar with the robot feel a sense of common ground when they see the happy expression, whereas newcomers may prefer interacting with the moody robot—

perhaps simply because a robot showing emotions is somewhat of a novelty. A social robot should therefore remember people who have interacted with it, and interact differently with those people than with newcomers. The robot should utilize its emotional expressions differently based on how much common ground it shares with a person.

Though we focused on the effect of the robot's mood, we found many factors that clearly contribute to how people interact with Valerie. Notable results include:

- Interactors who are in groups tend to interact among themselves as well as with the robot. This implies a need for the robot to acknowledge groups and present a means of engaging and interacting with multiple people at the same time.
- The time of day is consistently a factor in interaction patterns, with people who visit the robot in the afternoon tending to interact for the shortest periods of time. The robot should respond differently during different times of day—perhaps using shorter sentences or a faster speaking rate during times when people normally interact for short periods.

This research shows that only slight changes in a robot's expressions—and no changes in the structure of the interactions—influence how people perceive and interact with the robot. Different expressions created different expectations of the robot, with many people avoiding the negative robot or feeling more common ground with the positive robot. By applying these findings, we believe that we can greatly improve the quality of human-robot interaction. In addition, we believe that this work can stand as a basis for developing more sophisticated models of emotion for social robots.

7. ACKNOWLEDGMENTS

This work was funded in part by an NSF Graduate Research Fellowship to the first author and by NSF grants #IIS-032914 and #IIS-0121426. Additionally, we would like to thank Adam Kramer and Sara Kiesler for their assistance in the analyses presented in this paper.

8. REFERENCES

- [1] C. Breazeal. Emotion and sociable humanoid robots. *International Journal of Human Computer Studies*, 59:119–155, 2003.
- [2] A. Bruce, *et al.* The role of expressiveness and attention in human-robot interaction. In *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA)*, May 2002.
- [3] W. Burgard, *et al.* Experiences with an interactive museum tour-guide robot. *Artificial Intelligence*, 114(1-2):3–55, 1999.
- [4] J. Cassell, *et al.*, eds. *Embodied Conversational Agents*. MIT Press, 2000. ISBN 0262032783.
- [5] H. H. Clark. *Using Language*. Cambridge University Press, Cambridge, 1996.
- [6] C. Elliott. *The Affective Reasoner: A process model of emotions in a multi-agent system*. Ph.D. thesis, Northwestern University, 1992.
- [7] J. P. Forgas. Feeling and speaking: Mood effects on verbal communication strategies. *Personality and Social Psychology Bulletin*, 25(7):850–863, Jul 1999.
- [8] R. Gockley, *et al.* Designing robots for long-term social interaction. In *Proceedings of IROS 2005*. Edmonton, Alberta, 2005.
- [9] M. Montemerlo, *et al.* Experiences with a mobile robotic guide for the elderly. In *Proceedings of the National Conference of Artificial Intelligence (AAAI 02)*, pp. 587–592. Edmonton, AB, July 2002.
- [10] L. Moshkina and R. C. Arkin. On TAMEing robots. In *IEEE International Conference on Systems, Man and Cybernetics*, Oct 2003.
- [11] I. Nourbakhsh, *et al.* An affective mobile robot educator with a full-time job. *Artificial Intelligence*, 114(1-2):95–124, 1999.
- [12] B. Parkinson. Emotions are social. *The British Psychological Society*, 87:663–683, 1996.
- [13] A. Powers, *et al.* Common ground in dialogue with a gendered humanoid robot. In *Proceedings of RO-MAN 2005*. Nashville, TN, Aug 2005.
- [14] B. Reeves and C. Nass. *The Media Equation*. CSLI Publications, Cambridge, 1996.
- [15] W. S. Reilly and J. Bates. Building emotional agents. Tech. Rep. CMU-CS-92-143, School of Computer Science, Carnegie Mellon University, Pittsburgh, PA, 1992.
- [16] K. S. Rook, *et al.* When are dysphoric individuals distressing to others and vice versa? Effects of friendship, similarity, and interaction task. *Journal of Personality and Social Psychology*, 67(3):548–559, 1994.
- [17] R. Simmons, *et al.* GRACE: An autonomous robot for the AAI robot challenge. *AAAI Magazine*, 24(2):51–72, Summer 2003.
- [18] S. S. Sundar. Loyalty to computer terminals: is it anthropomorphism or consistency? *Behaviour and Information Technology*, 23(2):107–118, March-April 2004.
- [19] B. Wild, *et al.* Are emotions contagious? Evoked emotions while viewing emotionally expressive faces: quality, quantity, time course and gender differences. *Psychiatry Research*, 102(2):109–124, 2001.