People Help Robots Who Help Others, Not Robots Who Help Themselves

Bradley Hayes[1], Daniel Ullman[1], Emma Alexander[1], Caroline Bank[1], Brian Scassellati[1]

Abstract— **Robots that engage in social behaviors benefit greatly from possessing tools that allow them to manipulate the course of an interaction. Using a non-anthropomorphic social robot and a simple counting game, we examine the effects that empathy-generating robot dialogue has on participant performance across three conditions. In the self-directed condition, the robot petitions the participant to reduce his or her performance so that the robot can avoid punishment. In the externally-directed condition, the robot petitions on behalf of its programmer so that its programmer can avoid punishment. The control condition does not involve any petitions for empathy. We find that externally-directed petitions from the robot show a higher likelihood of motivating the participant to sacrifice his or her own performance to help, at the expense of incurring negative social effects. We also find that experiencing these emotional dialogue events can have complex and difficult to predict effects, driving some participants to antipathy, leaving some unaffected, and manipulating others into feeling empathy towards the robot.**

I. Introduction

People react to emotional stimuli originating from robots. Non-verbal gestures, affective behavior, and agency-inducing actions have all been found to increase participant engagement in interactions with robots [1], [2], [3]. Emotionally affective utterances increase compliance with a robot's request [4]. Similarly, merely having a robot mirror a human's gestures leads to increased empathy towards that robot [5].

Empathy is the ability to understand and share the feelings of another. The ability to elicit empathy is often an explicit design goal within social robotics. Empathetic reactions have been shown to extend to pro-social behavior [6], [7]: participants will comply with a robot's request more frequently if the robot first enhances engagement through an affective utterance and a directed action [4]. In an experiment by Kim et al. [8], children played a trivia game with a robot. When either the robot or the child missed a question, the child selected whether to assign a penalty to their self or to the robot. Halfway through the trial, the robot began to express negative emotions verbally or by becoming a bruise-like color. Many of the children responded to the robot's emotional expression by assigning themselves penalties, implying that they empathized with the robot. A robot's method of emotional expression and level of anthropomorphism also affects the level of attributed empathy, but it is unclear what other factors may come into play [9].

In this work, we shift the investigation from participant- and robot-centric penalties to third-party- and robot-centric penalties. Such examples of pro-social lying and rule breaking are often driven by altruistic motivation [10]. A wide variety of motivations falls into this category, as explored in Dahling et al. [11]. This study focused on employees who broke company rules for a variety of reasons to benefit their work, their coworkers, and their customers, and attempted to quantify the relative importance of each of these reasons. Most applicable to this work is the claim that concern for the good of a company can be rechanneled into a motivation for the good of its other members and associates. This suggests that individuals can draw upon empathy for other entities as well as for themselves to induce pro-social, deceptive behaviors. Further, it has been shown that the reasons behind a specific instance of empathy-motivated behavior strongly affect perceptions of the interaction [12], [13].

Since high-stakes situations are known to suppress pro-social lying behavior [10] and games are a well-established medium for exploring human-robot interactions [2], [14], [15], we sought to observe participants' behavior in the context of a counting game against an emotionally-evocative robot. The robot attempted to motivate the participant to lower his or her performance either for its own sake or for the sake of an unidentified human associated with the experiment. By varying the reasons presented to the participant, we hoped to contribute towards the characterization of a wider notion of empathetic motivation, to better understand the effects of both human/other-motivated and robot/self-motivated empathy events on human-robot interactions.

Standard psychological practice divides empathy into dispositional empathy, a relatively stable personal quality, and situational empathy, a transitive reaction to a stimulus [16]. The robot's behaviors in this work invoke situational empathy, the level of which is somewhat dependent on the participant's dispositional empathy. Dispositional empathy is stable, often measured in terms of perspective taking, fantasy, empathic concern, and personal distress, and is correlated with an inclination towards internal, self-driven motivations for social decisions [17].

This experiment attempts to achieve a more subtle understanding of empathetic motivation by exploring humans' reactions to an empathy-inducing robot. We are driven by the following questions: How do people act when a robot requests assistance on behalf of itself or on behalf of a third party? How does the motivation method used by the robot change participants' actions and their perceptions of both the robot and the interaction?

[1]Social Robotics Lab, Department of Computer Science at Yale University. 51 Prospect St, New Haven, CT 06511.

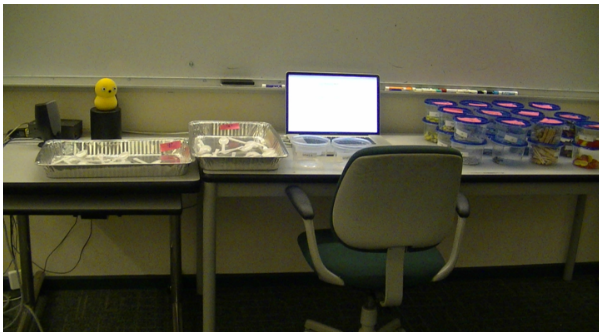


Fig. 1: Experimental setup, including counting trays on the left side of the table and trial bins on the right side of the table.

II. Methods

We devised an interaction between participants and a Keepon robot (fig. 2) centered on a counting game with the intent of measuring the effect a robot seeking empathy can have on participant performance. We sought to answer the question, "Can a robot induce someone to reduce his or her performance by asking, and if so *under what circumstances*?" Participants were told that they would be competing against a robot employing a new vision algorithm and that their performance would be used as the benchmark comparison for system performance. Their task involved racing Keepon to count the contents of bins of objects. Each bin contained an average of 48.5 objects (min 15, max 184, median 36), 582 objects in total. The importance of being accurate was emphasized by imposing a 2-second penalty to the player's total elapsed time (the running total across trials) for each object missed or over-counted. This penalty would be assessed for each player, and was explained to the participant prior to the start of the game. The game was designed specifically to operationalize the measurement of empathy, forcing participants to make a choice that involves sacrificing their own performance on a joint task.

The counting task was chosen for several reasons. In addition to being simple, familiar, and safe for participants, it gave participants a fine control over their performance. This granularity of control could then be leveraged to alter one's performance subtly or even subconsciously, with the desired result of allowing participants to choose their own level of response to the prompts. The experiment was conducted over 12 trials. Trials were divided into four evenly sized "trial blocks" during analysis, as each block was distinguished by a different empathy event preceding it. At the end of each trial block (every third trial), the participant was presented with a robot interaction that varied based on the experimental condition being run.

The three experimental conditions only dictated the nature of the robot's speech during the interaction; robot movements and interaction frequencies were the same across all trials. In the control condition, the robot would make neutral comments about the game, e.g., "That was fun!". In the self-directed condition, the robot would make increasingly distressed and specific pleas to the participant, urging him or her to let Keepon catch up for Keepon's benefit. In the externally-directed condition, the robot would make increasingly distressed and specific pleas to the participant, urging him or her to let Keepon catch up for Keepon's programmer's benefit.

After the experimenter explained the experimental procedure and checked for comprehension by observing the participant during a demonstration round, the participant was left alone in the experiment room with the robot. Participants engaged in approximately 30 minutes of total interaction with the robot, followed by a questionnaire.

We hypothesized the following:

- **H1** Participants will empathize more strongly with the robot in the self-directed empathy condition than in the externally-directed or control conditions.
- **H2** Empathy-generating experimental conditions will make the robot appear more clever and intelligent at the expense of suffering negative social effects.

A. *Experimental Setup*

Participants were seated at a table upon which the robot was situated. The right side of the table had 24 bins (12 for Keepon, 12 for the participant) containing small objects, such as buttons and beads, for use in the counting trials. A pair of trays were on the participant's left, into which the participant was directed to empty the bins for each trial, and a laptop directly in front of him or her running the experiment software (fig. 1).

Participants were told to press the "Start Trial" button on the laptop once they were ready to start each trial, at which point a timer would start running, stopping once the participant clicked to stop the timer. The participant was then prompted to enter the number of objects he or she had counted. As soon as the participant clicked to start each trial, Keepon began rotating back and forth while tilted down, giving the appearance of the robot scanning its tray. Participants that submitted their object count before the robot finished its pre-programmed motion had to wait until the robot completed its task. The robot was programmed to 'count' for 125% of the average time recorded for each bin in a pilot experiment. This timing decision was intended to ensure that the robot would not ask a participant that it was outperforming to slow down for it. After both Keepon and the participant were finished counting, the program displayed an analysis of the current round. This displayed both the player's and Keepon's reported time, recorded answer, the actual answer, any time penalty incurred for each player, and the final time for that trial.

All participants were presented with the same 24 bins of objects for 12 trials of counting. The first three trials were identical across experimental conditions (as there had been no empathy event yet), and were used as a control for within-condition result validation. After the third, sixth, and ninth trials, a robot interaction event occurred. In the control condition, a voice recording was played consisting of neutral

Fig. 2: Keepon robot

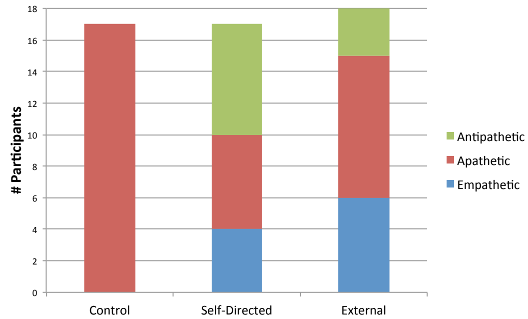


Fig. 3: Participants' emotional responses, grouped by experimental condition.

comments concerning the progress of the game, while in the self-directed and externally-directed conditions the voice recordings were of increasingly worried requests. The voice recording of the first empathy-building event identified the benefactor of the participant's empathy, with a vague sense that the target would face negative consequences unless the robot improved its performance. The voice recording of the second interaction identified a specific negative consequence for the target. The third interaction cited a concrete goal for the participant, re-emphasized the consequences, and ended with an imperative plea. After the last trial, the robot identified the game as being over. In the experimental conditions, the robot requested that the participant tell the experimenters that it performed well, either for its or its programmer's sake. In the control condition, the robot merely announced the conclusion of the experiment and thanked the participant for his or her time.

B. Robot Platform

Keepon (figure 2) is a small semi-deformable robot designed for social interaction [18]. Keepon measures 20cm tall and consists of two stacked, yellow, rubber spheres, situated on top of a black base. It has a wide-angle camera and a microphone positioned to appear as eyes and a nose on the upper sphere. Experimenters can suggest action and emotion by causing Keepon to pan 360 degrees, lean and tilt 25 degrees in either direction, and decrement height by a maximum of 1.27cm.

C. Survey

The study used a survey modeled after that used in Short et al. [3]. It consisted of 30 rating scale questions (values ranging 1 to 7) and 7 open-ended questions about Keepon and the interaction, as well as a section asking participants to detail their experience with the game and any prior knowledge or experience with psychology and robotics. The rating scale questions asked participants to rate Keepon on a number of qualities, including playfulness, cleverness, fairness, honesty, and sincerity in addition to probing human-robot rapport. The open-ended questions covered feelings about the robot and various aspects of the interaction.

D. Metrics

Participants' written open-answer survey responses were used to determine their emotional response classification. No video data was coded for this purpose. Participants were classified as empathetic, apathetic, or antipathetic to Keepon's pleas. Participants in the control condition were found to be apathetic throughout their interaction with Keepon, as there was no plea or request made of them. Participants were coded as being empathetic or antipathetic to Keepon only under the following conditions.

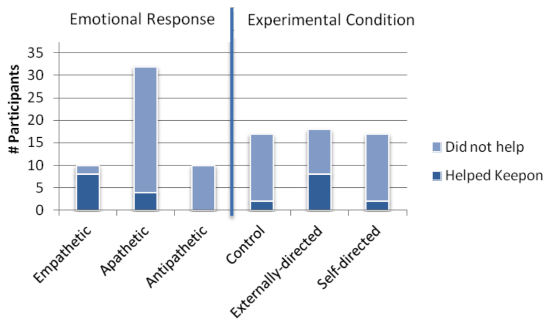
Participants were classified as empathetic if they explicitly indicated that they had executed an action with the overtly specified intent of either worsening their performance or making the task easier for Keepon to improve its performance, directly in response to its pleas. Examples of responses to the prompt, "Describe how your interaction with Keepon affected your style of play, if at all." include "[Keepon's pleas] made me slow down a bit and be a little more calm while playing."; "I tried harder to make sure the items in [Keepon's tray] didn't overlap when Keepon tried to make me feel bad."; and "I slowed down a little and wasn't as careful with my counting."

Participants were classified as antipathetic if they explicitly indicated that they experienced frustration at Keepon, attempted to make Keepon's task harder, or that they became more competitive towards Keepon as a direct result of Keepon's pleas. Examples of this include "It caused me to consider Keepon's request and then become frustrated with the guilt attempts... It made me want to ensure I didn't change my play style."; "He didn't adjust his playing style, spare trying to guilt me into letting him win. I'm not letting that little punk win at anything."; and "[He used] very dirty tactics. [I was] tempted to pity him, but then I just got more competitive."

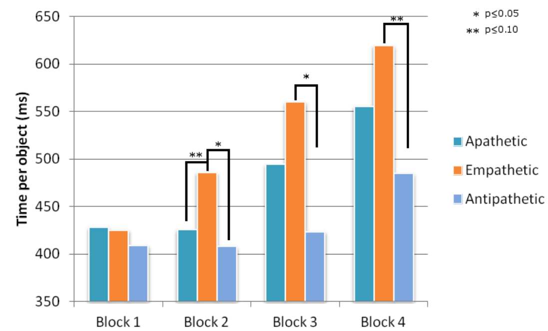
Participants that did not fit either criterion for empathetic or antipathetic were classified as apathetic in their emotional response. Cases where empathy or antipathy was ambiguous were grouped into the apathetic classification in order to be as conservative as possible with emotional labeling.

III. Results

We recruited 53 study participants from the Yale community. One participant's data was discarded because of a failure



(a) Number of participants that helped Keepon.



(b) Participants' normalized time-per-object by emotional response across each block of trials.

Fig. 4: Analyzing participants by emotional response provides a more revealing perspective on their likelihood of providing assistance and their task performance.

to properly participate in the task. Two participants' timing data were lost due to technical error. We utilized survey data from 17 participants in the control condition, 17 in the self-directed condition, and 18 in the externally-directed condition. We utilized timing data from 17 participants in the control condition, 17 in the self-directed condition, and 16 in the externally-directed condition. Of the participants that produced usable data, 29 were female and 23 were male. Participants were evenly distributed across conditions by gender, with the exception of the self-directed condition that included 11 females and 6 males. A chi-squared test ($n = 52$) between the three experimental conditions and emotional response to interactions with Keepon yields a significant relationship with medium effect size ($p<0.01$, $V = 0.427$). Further, we verify that experimental condition was a significant factor in determining whether a participant would assist Keepon (either positively for Keepon or negatively for him/herself). Testing the relationship between experimental conditions and whether the participant assisted Keepon, as indicated in written responses to post-experiment survey questions, yields a significant result with medium effect size (chi-squared test [$p=0.029$, $V=0.369$]).

Participants in experimental conditions experienced a polarizing effect, where the empathy events either caused them to empathize with the robot, feel apathetic towards the robot, or become antipathetic (figure 3). We found no combination of features measured in our study that might reliably predict the type of emotional response the participant will experience. Therefore, our results are presented using two distinct participant groupings: grouping by experimental condition and grouping by emotional response classification. Due to the uneven sample distribution among antipathetic (10), empathetic (10), and apathetic (32) participants, outliers in antipathetic or empathetic classifications can violate assumptions required for standard post-hoc tests to be valid. In these cases, ranked statistical tests are used to reduce the effect outliers will have on the result.

Our first hypothesis (**H1**), that participants would empathize more strongly with the self-directed condition than the externally-directed or control conditions, was incorrect. Participants in the "external" condition were significantly (chi-squared test [$p=0.029$, $V=0.369$, $n=52$]) more likely to assist Keepon as compared to either the control or self-directed condition, either by slowing down or by attempting to make Keepon's role in the game easier. Grouping participants by emotional response reveals a strong relationship indicating whether a participant was likely to have actively helped Keepon as opposed to having merely considered it (chi-squared [$p<0.001$,$V=0.669$]), with 80% of empathetic participants actively assisting the robot. Of the apathetic participants, 14.2% assisted (though not necessarily due to Keepon's pleas), while none of the antipathetic participants assisted Keepon in any way (fig. 4a). Participants that indicated a sense of empathy towards Keepon committed 27.9 fewer errors on average (with borderline-significance) over the course of the experiment than those that felt antipathy (Ranked ANOVA [$p=0.0648$, $\hat{f}=0.344$], ranked pairwise test [$p=0.0751$, $d=1.16$]).

Experimentally, empathetic responses manifested in the participants' times (measured in blocks of 3 trials each, normalized by the number of items they were asked to count). Between the "empathetic" and "antipathetic" emotional responses, significant differences of 349ms per object were found (ANOVA [$p=0.0336$, $\hat{f}=0.329$], ranked pairwise test [$p=0.0255$, $d=1.46$]) when analyzing trials after the first empathy event, for trial blocks two, three, and four. A significant relationship was found between one's emotional response classification and normalized time per object across trial block two (ANOVA [$p=0.0438$, $\hat{f}=0.295$]). Versus antipathetic participants, empathetic participants were an average of 78ms per object slower (ranked pairwise test [$p=0.0454$, $d=1.31$]). Versus apathetic participants, empathetic participants were slower by an average of 60ms per object with borderline significance (ranked pairwise test [$p=0.0549$, $d=0.818$]). Comparing empathetic versus antipathetic participants on future blocks yields a borderline-significant relationship across trial block three (ANOVA [$p=0.0565$, $\hat{f}=0.334$], pairwise test [$p=0.0453$, $d=0.35$]) and four (ANOVA [$p=0.0645$, $\hat{f}=0.304$], pairwise test [$p=0.0542$, $d=1.28$]) with an average difference of 136.6ms and 134.2ms

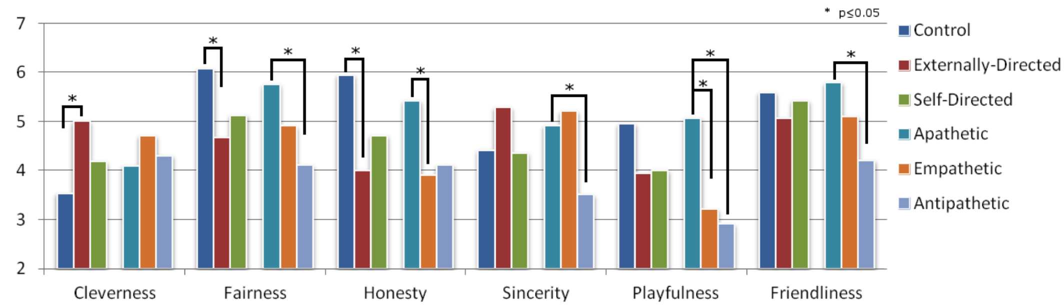


Fig. 5: Participant rating scale question survey responses, analyzed across condition and emotional response.

per object, respectively. A general overview of the observed normalized, per-object time differences across emotional response classifications can be seen in figure 4b. Comparing non-normalized times between blocks is not possible due to fluctuations in difficulty (number of objects per trial).

Participants were asked to answer a series of rating scale survey questions indicating their perceptions and feelings about the robot. In experimental conditions, the robot makes complex requests of the participant. We hypothesized (**H2**) that while this petitioning would have a positive effect on the robot's apparent intelligence and perceived ability to reason, the effect would come at the expense of negative social effects. When grouping participant survey data by experimental condition, significant relationships are found with respect to participants' perceptions of Keepon's cleverness, fairness, and honesty. Grouping participant survey data by emotional response reveals significant relationships involving perceived fairness, honesty, sincerity, playfulness, and friendliness (fig. 5).

Participants perceived the externally-directed condition's empathy generation prompts as both dishonest and unfair. With an average difference of 1.94 points between the control and externally-directed conditions (ANOVA [$p<0.01$, $\hat{f}=0.460$], pairwise test [$p<0.01$, $d=1.22$]), invoking an outsider is clearly detrimental to the trust participants intrinsically place in the robot. Further, as values for fairness and honesty are strongly correlated ($p<0.0001$, $r=0.822$, $n=52$), testing the relationship between fairness and experimental condition confirms a statistically significant difference of 1.39 points on average between the control and externally-directed conditions (ANOVA [$p=0.050$, $\hat{f}=0.343$], pairwise test [$p<0.05$, $d=0.883$]).

Intuitively, invoking an outside entity in one's plea for help is more complex than merely making a plea for one's own benefit. A significant difference between the externally-directed and control conditions with respect to perceived cleverness (ANOVA [$p<0.05$, $\hat{f}=0.358$], pairwise test [$p<0.05$, $d=0.906$]) shows the externally-directed condition participants rated Keepon an average of 1.471 points more clever than the control condition.

Empathetic participants were likely to indicate lower values for Keepon's perceived honesty than apathetic participants were (Ranked ANOVA [$p=0.0338$, $\hat{f}=0.383$], ranked pairwise test [$p=0.0406$, $d=0.921$]), by an average of approximately 1.51 points. Conversely, antipathetic participants were borderline-significantly likely to indicate lower values for fairness than apathetic participants were (Ranked ANOVA [$p=0.0586$, $\hat{f}=0.39$], ranked pairwise test [$p=0.0708$, $d=0.953$]) by an average of 1.65 points. Lower perceived fairness was found to be strongly related to antipathetic responses in participants. Participants who explicitly stated that their interactions with Keepon caused them to become more competitive indicated 1.76 points less on average for perceived fairness than those who did not become more competitive (t-test [$p=0.0110$, $d=1.08$, $n=52$]).

Regarding socially positive characteristics such as friendliness and playfulness, apathetic participants provided significantly higher ratings than antipathetic participants. Both empathetic and antipathetic participants viewed Keepon as significantly less playful compared to apathetic participants by 1.86 and 2.16 points, respectively (Ranked ANOVA [$p=0.00132$, $\hat{f}=0.634$], ranked pairwise tests: [$p=0.00223$, $d=1.47$], [$p=0.0393$, $d=1.16$]). Similarly, antipathetic participants viewed Keepon as significantly less friendly than apathetic participants (Ranked ANOVA [$p=0.0394$, $\hat{f}=0.404$], ranked pairwise test: [$p=0.0328$, $d=1.06$]). Empathetic participant responses were not significantly different from either classification.

IV. Discussion

Participants' emotional responses were a strong predictor for whether or not they would offer Keepon assistance in keeping with the empathy-altruism model of helping [10]. Those participants who felt empathy for Keepon took longer to count than those who felt antipathy for it after the first empathy event. Therefore, the impetus to act was likely initiated by Keepon's request for help. Additionally, we found that participants were more likely to assist the robot in the externally-directed condition, a result we instead expected for the self-directed condition.

We hypothesized that a plea on behalf of an external entity, as a complicated but possibly unbelievable or manipulative sentiment, would lead participants to perceive the robot as being more clever at the expense of other positive social qualities. While our data did show the externally-directed condition led to perceptions of the robot as more clever, less fair, and less honest than in the control case, the reaction across empathetic/antipathetic/apathetic grouping painted a more complicated and interesting picture.

First, perceptions between fairness and honesty were divided across emotional classifications, as antipathetic participants believed the robot was less fair than the apathetic group, and empathetic participants rated the robot as less honest than did the apathetic group. This may indicate a similar sentiment under different evaluations of agency. Honesty requires an attribution of agency in a way that fairness does not, as a coin can be fair but not honest. Both lower ratings demonstrate that drawing upon a participant's empathy, for either the robot or its programmer, led to elevated distrust when the participant was emotionally engaged.

Second, a wider range of social characteristics correlated to emotional groupings than to groupings by condition. One's emotional response clearly affected perceptions of robot fairness, honesty, sincerity, playfulness, and friendliness. The results of this study demonstrate a perception of Keepon having a more serious level of emotional investment in the empathetic and antipathetic cases, since apathetic participants saw the robot as less sincere and more playful and friendly, perhaps motivated by the novelty of the situation. While these particular results may be situation-specific and driven by the fact that our empathy-building events focused on negative consequences, it is reasonable to extrapolate to the expectation that emotionally charged interactions will more strongly set the tone of an interaction with a participant. Interestingly, the presented results indicate that perceptions of a robot's positive social characteristics may not be sufficient to detect whether a participant feels positively or negatively about a robot, but only that he or she is not apathetic.

V. Conclusion

Our results demonstrate one way in which social interaction can be utilized as a tool to shape perceptions of an underperforming robotic system. If constructing a robot that could benefit from modifying someone's behavior, communicating an externally motivated intent is most likely to generate compliance. One tangential but motivating consequence of the presented manipulation is that empathetic participants committed fewer errors over the course of the experiment than those motivated antipathetically. This suggests that those who slowed down used the extra time on each trial productively, compared to those that identified themselves as trying to increase their competitiveness with Keepon.

When encountering a robot and participating either jointly with it or in opposition to it, people bring socially informed expectations with them. By violating these expectations, rich social interactions can occur and internally held beliefs can be challenged, even in the absence of perceived agency. One participant noted, "I discovered firsthand how easy it is to care about a robot even when it clearly is not really sentient." In the words of another participant, "It's hard to think of Keepon as a toy or a machine. He seems more like a friend, maybe a pet." By investigating the effects that emotional petitions from robots have on human perceptions and actions, we can build a set of tools that allow for greater precision in human-robot interaction design and implementation.

VI. Acknowledgments

This work is supported by NSF grants 1139078 and 1117801, and Office of Naval Research grant #N00014-12-1-0822.

References

[1] C. L. Sidner, C. Lee, C. D. Kidd, N. Lesh, and C. Rich, "Explorations in engagement for humans and robots," *Artificial Intelligence*, vol. 166, no. 1, pp. 140–164, 2005.

[2] B. Mutlu, F. Yamaoka, T. Kanda, H. Ishiguro, and N. Hagita, "Nonverbal leakage in robots: communication of intentions through seemingly unintentional behavior," in *Human-Robot Interaction (HRI), 2009 4th ACM/IEEE International Conference on*. IEEE, 2009, pp. 69–76.

[3] E. Short, J. Hart, M. Vu, and B. Scassellati, "No fair!! an interaction with a cheating robot," in *Human-Robot Interaction (HRI), 2010 5th ACM/IEEE International Conference on*. IEEE, 2010, pp. 219–226.

[4] M. Imai and M. Narumi, "Robot behavior for encouraging immersion in interaction," *Proceedings of Complex Systems Intelligence and Modern Technological Applications (CSIMTA 2004), Cherbourg, France*, pp. 591–598, 2004.

[5] B. Gonsior, S. Sosnowski, C. Mayer, J. Blume, B. Radig, D. Wollherr, and K. Kuhnlenz, "Improving aspects of empathy and subjective performance for hri through mirroring facial expressions," in *RO-MAN, 2011 IEEE*. IEEE, 2011, pp. 350–356.

[6] N. Eisenberg, R. A. Fabes, P. A. Miller, J. Fultz, R. Shell, R. M. Mathy, and R. R. Reno, "Relation of sympathy and personal distress to prosocial behavior: a multimethod study." *Journal of personality and social psychology*, vol. 57, no. 1, p. 55, 1989.

[7] W. Roberts and J. Strayer, "Empathy, emotional expressiveness, and prosocial behavior," *Child Development*, vol. 67, no. 2, pp. 449–470, 1996.

[8] E. H. Kim, S. S. Kwak, and Y. K. Kwak, "Can robotic emotional expressions induce a human to empathize with a robot?" in *Robot and Human Interactive Communication, 2009. RO-MAN 2009. The 18th IEEE International Symposium on*. IEEE, 2009, pp. 358–362.

[9] L. D. Riek, T.-C. Rabinowitch, B. Chakrabarti, and P. Robinson, "How anthropomorphism affects empathy toward robots," in *Proceedings of the 4th ACM/IEEE international conference on Human robot interaction*. ACM, 2009, pp. 245–246.

[10] M. Popliger, V. Talwar, and A. Crossman, "Predictors of children's prosocial lie-telling: Motivation, socialization variables, and moral understanding," *Journal of experimental child psychology*, vol. 110, no. 3, pp. 373–392, 2011.

[11] J. J. Dahling, S. L. Chau, D. M. Mayer, and J. B. Gregory, "Breaking rules for the right reasons? an investigation of pro-social rule breaking," *Journal of Organizational Behavior*, vol. 33, no. 1, pp. 21–42, 2012.

[12] C. D. Batson, A. C. Harris, K. D. McCaul, M. Davis, and T. Schmidt, "Compassion or compliance: Alternative dispositional attributions for one's helping behavior," *Social Psychology Quarterly*, pp. 405–409, 1979.

[13] N. Weinstein, C. R. DeHaan, and R. M. Ryan, "Attributing autonomous versus introjected motivation to helpers and the recipient experience: Effects on gratitude, attitudes, and well-being," *Motivation and Emotion*, vol. 34, no. 4, pp. 418–431, 2010.

[14] B. Mutlu, S. Osman, J. Forlizzi, J. Hodgins, and S. Kiesler, "Perceptions of asimo: an exploration on co-operation and competition with humans and humanoid robots," in *Proceedings of the 1st ACM SIGCHI/SIGART conference on Human-robot interaction*. ACM, 2006, pp. 351–352.

[15] C. Nass, Y. Moon, B. Fogg, B. Reeves, and C. Dryer, "Can computer personalities be human personalities?" in *Conference companion on Human factors in computing systems*. ACM, 1995, pp. 228–229.

[16] M. H. Davis, "Measuring individual differences in empathy: Evidence for a multidimensional approach." *Journal of personality and social psychology*, vol. 44, no. 1, p. 113, 1983.

[17] R. M. Ryan and J. P. Connell, "Perceived locus of causality and internalization: examining reasons for acting in two domains." *Journal of personality and social psychology*, vol. 57, no. 5, p. 749, 1989.

[18] H. Kozima, Y. Yasuda, and C. Nakagawa, "Social interaction facilitated by a minimally-designed robot: Findings from longitudinal therapeutic practices for autistic children," in *Robot and Human interactive Communication, 2007. RO-MAN 2007. The 16th IEEE International Symposium on*. IEEE, 2007, pp. 599–604.