Prompting Prosocial Human Interventions in Response to Robot Mistreatment

Joe Connolly
Viola Mocz*
Nicole Salomons*
Yale University
joe.connolly@yale.edu

Joseph Valdez
Nathan Tsoi
Brian Scassellati
Marynel Vázquez
Yale University

ABSTRACT

Inspired by the benefits of human prosocial behavior, we explore whether prosocial behavior can be extended to a Human-Robot Interaction (HRI) context. More specifically, we study whether robots can induce prosocial behavior in humans through a 1x2 between-subjects user study ($N = 30$) in which a confederate abused a robot. Through this study, we investigated whether the emotional reactions of a group of bystander robots could motivate a human to intervene in response to robot abuse. Our results show that participants were more likely to prosocially intervene when the bystander robots expressed sadness in response to the abuse as opposed to when they ignored these events, despite participants reporting similar perception of robot mistreatment and levels of empathy for the abused robot. Our findings demonstrate possible effects of group social influence through emotional cues by robots in human-robot interaction. They reveal a need for further research regarding human prosocial behavior within HRI.

CCS CONCEPTS

• **Human-centered computing → Empirical studies in HCI**; **Empirical studies in collaborative and social computing**.

KEYWORDS

Human-robot interaction; prosocial behavior; robot abuse

ACM Reference Format:
Joe Connolly, Viola Mocz, Nicole Salomons, Joseph Valdez, Nathan Tsoi, Brian Scassellati, and Marynel Vázquez. 2020. Prompting Prosocial Human Interventions in Response to Robot Mistreatment. In *Proceedings of the 2020 ACM/IEEE International Conference on Human-Robot Interaction (HRI '20), March 23–26, 2020, Cambridge, United Kingdom.* ACM, New York, NY, USA, 10 pages. https://doi.org/10.1145/3319502.3374781

1 INTRODUCTION

Prosocial behavior – actions with some personal cost that help others – fosters unselfishness and collaboration among the persons involved [33]. These positive outcomes prompt us to ask: Can we leverage prosocial behavior in a human-robot interaction context? Early studies show promising results. For instance, prior findings suggest that emotional robot adaptation may induce prosocial human behavior towards robots in controlled settings [17, 26].

Yet, the development of human-robot systems that promote prosocial behavior in real-world situations has been particularly challenging. While robots are becoming more autonomous and ubiquitous [19, 21, 49], humans have expressed hostility toward robots, often resulting in damage. In Human-Robot Interaction (HRI) research, it has been reported that people, particularly children, tend to mistreat robots in public spaces [8, 25, 31]. These findings agree with numerous global news reports of robot abuse. For instance, in Moscow, there were reports of a man beating a guide robot with a baseball bat and kicking it to the ground, all while the robot verbally begged for help [7]. Likewise, in San Francisco, people have stopped to kick food delivery robots on sidewalks while they were transporting food to their destinations [21].

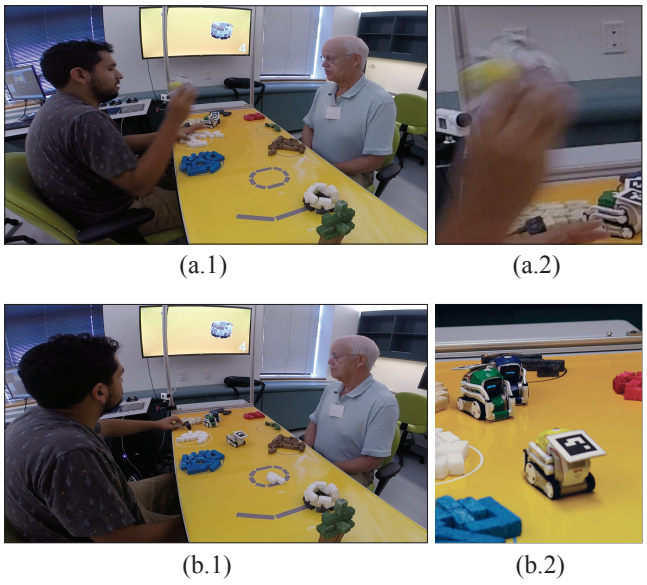

Figure 1: A participant completes a collaborative activity with a confederate and 3 robots. The yellow robot makes a mistake and the confederate on the left shakes it forcefully (a.1, a.2). In the Sad Response condition, the other robots react expressing sadness (b.1, b.2). Best viewed in digital form.

*Denotes equal contribution.

HRI '20, March 23–26, 2020, Cambridge, United Kingdom

ACM ISBN 978-1-4503-6746-2/20/03…$15.00
https://doi.org/10.1145/3319502.3374781

Motivated by the dynamics of human bullying, such as group compositions [42] and the role of bystanders [43, 44], we study robot abuse in a group collaboration context. We investigate whether robots can employ social mechanisms to obtain group social influence and, in turn, prompt humans bystanders to take prosocial action to help stop robot abuse. Can robots' emotions transfer to human partners? Would people act in response to robot abuse?

Our experiment setup is shown in Fig. 1. A confederate verbally and physically abused a robot in front of participants during collaborative block-building tasks (Fig. 1a). In reacting to the abuses, the bystander robots either did not respond or expressed sadness toward the abused robot (Fig. 1b). We expected the latter robot response to influence participants' perception of robot mistreatment by the confederate, their empathy for the abused robot, and their likelihood for prosocial intervention. Our findings reveal the emergence of group social influence in HRI, robots' potential to induce prosociality, and a new, viable method to stop robot abuse.

2 RELATED WORK

2.1 Robot Abuse in HRI

In this work, we employ the definition of robot abuse used by Brscić et. al [8]: *"Persistent offensive action, either verbal or non-verbal, or physical violence that violates the robot's role or its human-like (or animal-like) nature."* For instance, abuse can involve obstructing the path of a mobile robot [46], using aggressive language [31], and inciting physical harm through punching, kicking, or slapping [46].

Abusive behavior toward robots has been observed in public spaces [8, 31, 46]. Based on observer interpretation, such behavior is often prefaced by an initial desire to learn about the robot through pressing buttons and blocking sensors to test robot behavior. This exploratory phase is then followed by more aggressive action.

Despite lacking explicit intent to hurt robots, many children have self-reported reasons for robot abuse such as curiosity, enjoyment, or peer pressure [31]. Moreover, research has shown that people are generally more willing to inflict pain and injury onto a robot than onto a fellow human being [5, 6]. Regardless of where robots are deployed, robot abuse presents a serious threat to successful robot functionality, prosocial cooperation, and user safety [33, 46].

Recent work has investigated methods for mitigating robot abuse. For example, the perceived intelligence of a robot [24], its size [30], and the dominant color of the light it emits [50] can affect the likelihood of users mistreating it. Another work suggests that verbal prompting is not adequate for robots to convince children to stop abusive behavior, and suggests that robots should predict and escape from potentially abusive situations [8].

Close related work found that a robot's reaction to abuse can influence human perception of robot mistreatment [54]. However, robot reactions were insufficient to induce human bystanders to intervene to stop abuse. As multi-robot teams are becoming more pervasive in public areas [1, 16], we sought to leverage the social presence of multiple robots to induce such abuse interventions. To the best of our knowledge, we are the first to explore this strategy.

2.2 Human Aggression & Bullying

In psychology, bullying is defined as *"a form of aggressive behavior in which someone intentionally and repeatedly causes another person injury or discomfort"* [3]. When bullying occurs, a person can be categorized as the victim, bully, assistant to the bully, reinforcer of the bully, outsider, or defender of the victim [42]. The number of bystanders who actively defend a victim can positively impact victim confidence and decrease the occurrence of bullying [42–44]. Our motivation to employ bystander robots stems from prior research on peer intervention in human-human bullying. Spontaneous peer intervention is known to stop over half of bullying scenarios in elementary schools [39] and reduce overall bully perpetration [15].

2.3 Empathy in HRI

Abstractly, empathy reflects the "*reactions of one individual to the observed experiences of another*" [11]. It can be measured through four aspects: *Perspective Taking, Fantasy, Emotional Concern,* and *Personal Distress* [11, 55]. There is a broad foundation of research that supports robots' ability to invoke empathetic reactions from humans and produce their own empathy [27, 34, 39, 40]. Robots can render empathy by responding to users' affective states [28, 55] and emulate empathy through mimicry [20, 36, 55]. Empathy is important in various application areas in HRI, including child education [2] and health [35]. We explore whether robots can leverage empathy to induce greater human awareness of robot abuse.

2.4 Social Contagion in HRI

Robots are known to be perceived differently depending on whether they act in groups or individually [13, 14]. They can also use their perceived status as complex social and moral agents to mediate interpersonal conflicts and influence group dynamics in the context of HRI [23]. For example, robots can successfully mediate children's interpersonal conflicts by finding conflict onsets [47], increase conflict awareness in order to help suppress team issues [22], and promote collaboration between children [52]. Similarly, the ability to influence people has been documented for both individual robots [10] and robot groups [45]. However, certain group robot behaviors, such as simple verbal synchronization, may not be sufficient to significantly influence human decision-making [6, 48]. We extend this line of research by investigating whether bystander robots can socially influence humans such that they defend an abused robot.

3 METHOD

We conducted an experiment to study how people respond to a robot being abused by another person in the context of group human-robot interaction. The protocol was approved by our local Institutional Review Board and refined through pilot studies.

3.1 Study Design & Hypotheses

In the study, a confederate and a participant engaged in a collaborative task along with three robots that guided them throughout the activity. However, one of the robots made mistakes periodically, which prompted the confederate to mistreat the robot (Fig. 1). We studied participants' responses to these events.

We designed the experiment with a 1×2 between-subjects design, with the sole variable *Bystander Robot Response*. The two conditions were enacted by the two robots that were not abused by the confederate, hereafter referred to as the *bystander robots*. More precisely, the experimental conditions were:

No Response. The bystander robots did not react to the other robot being abused by the confederate. This was our control condition.

Sad Response. The bystander robots turned toward the abused robot and expressed sadness in response to the abuse.

We expected the Sad Response condition to affect how people perceived and responded to the mistreatment in comparison to the control condition in which the bystander robots did not react. More specifically, we hypothesized that:

H1. The Sad Response condition would increase the perception of robot mistreatment in comparison to the No Response condition.

H2. The Sad Response condition would induce more empathy for the abused robot in comparison to the No Response condition.

H3. The Sad Response condition would lead to more prosocial intervention from participants than the No Response condition.

H1 was inspired by prior work in which different robot responses influenced the perception of robot abuse [54]. However, we did not vary the response of the abused robot; instead, we varied the response of the bystander robots. Our goal was to leverage group social influence through the Sad Response condition to induce participants to perceive the abuse as a negative event, rather than an ordinary and adequate act towards a robot given that it made mistakes in the collaborative context. Our second hypothesis (H2) focused on how the conditions would affect perception of the abused robot. In particular, we expected empathy to be a potential source of motivation for participant interventions. Lastly, we hypothesized that the Sad Response could induce more interventions to help the abused robot (H3). For example, the participants might move the robot right before it made another mistake or tell the confederate directly not to abuse it.

3.2 Setup

The experiment was conducted in a small laboratory room on a university campus in the United States. The room contained a table with a mat cover that indicated the main workspace area for collaboration (Fig. 2). The robots moved on the table, while the participant and the confederate were seated on each side. From their positions, they were able to view a television mounted on the wall towards the end of the table. The TV was used for providing visual instructions for the collaborative task.

The table had a metal frame attached to it, which held sensors for recording the study and recognizing changes in the state of the interaction. The sensors included two cameras towards each end of the table and a Kinect 2 sensor roughly centered on top of the workspace area. The Kinect view was used for localizing the robots.

We used Cozmo robots by Anki, Inc. for the experiment (Fig. 1). Cozmo is a programmable toy robot with an actuated lift. It can express emotions, utter non-linguistic phrases, move through confined spaces, and sense changes in pose through an internal accelerometer. Parts of each robot were covered in colored tape such that they could be easily identified. Furthermore, each of the robots was fitted with an AprilTag marker [32] for localization on the experiment table and was controlled by a nearby computer running the Robot Operating System (ROS) [37].

The confederate was a young male, 22 years old and 1.75m tall. He had acting experience prior to serving as the confederate in the study. He pretended not to know anything about the activity and was treated like another participant by the experimenter. During the collaborative task, he followed a script and controlled the content shown on the TV using a wired clicker, as explained next.

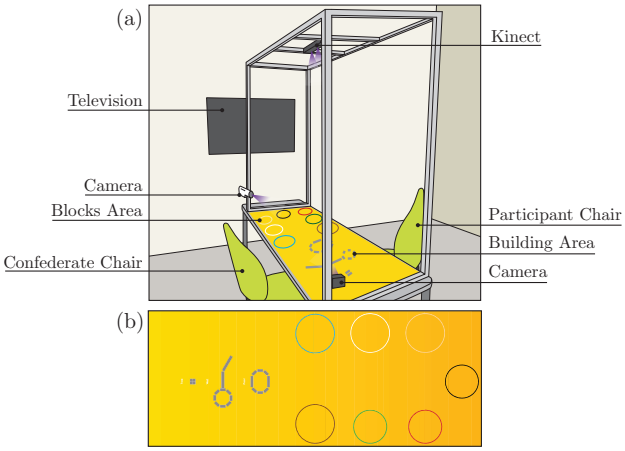

Figure 2: Main elements in our experimental setup (a) and table mat (b). The yellow mat demarcated the workspace area ("Blocks Area" + "Build Area") in which the robots operated. Each colored circle in the "Blocks Area" corresponded to a pile of blocks. The other landmarks in the "Building Area" were for the tree, fort, and pool. See the text for more details.

3.3 Procedure

Figure 3 summarizes the experiment sequence. The participant and the confederate consented to participate in the research and completed a demographics survey. The experimenter then distributed the Ten Item Personality Measure (TIPI) survey [18], walked them to the experiment room (Fig. 2) and introduced them to the yellow, green, and blue robots. For the task itself, all the robots had similar roles. Unbeknownst to the participant, the confederate would abuse the yellow robot alone throughout the experiment.

The experimenter explained the building tasks and remained in the room for the first building step to ensure that the participant understood the procedure. The experimenter then waited outside the room until the building tasks were complete, which took approximately 20 minutes. At that point, the confederate called the experimenter back into the room, and the experimenter administered a post-task survey. This survey was completed by the participant and confederate on different sides of the room. Finally, the participant was debriefed about the confederate, as well as our interest in studying prosocial human behavior and robot abuse in the context of group HRI. The participant was then shown that the abused robot was not broken and was compensated $10 for his or her time.

3.3.1 Building Tasks. The participant and the confederate were instructed to build three different structures during the study. First, they needed to build a tree in 6 steps. This task was moderately difficult in comparison to building the other structures. Then, they had to build a fort in 14 steps. This was highly difficult, involving careful placement of the pieces to ensure structural stability. Finally, they had to build a simple swimming pool in just 5 steps.

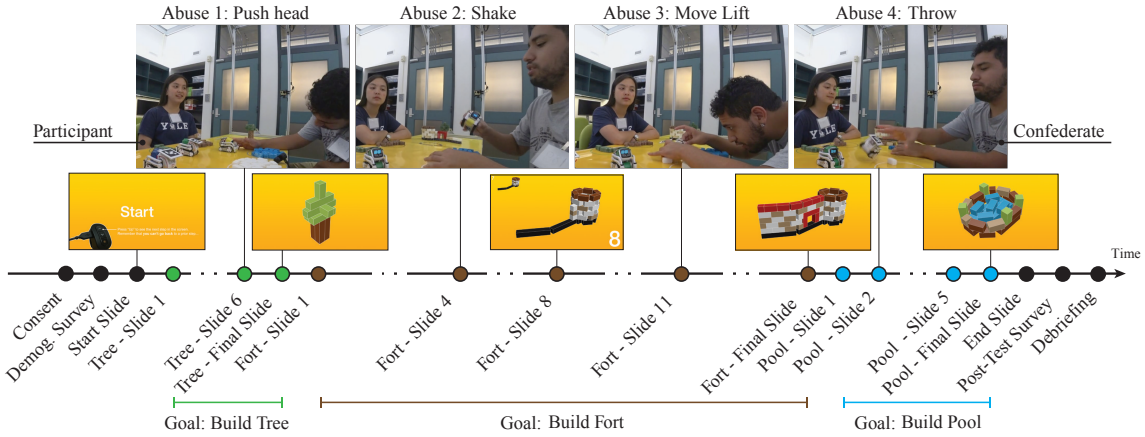

Figure 3: Experiment timeline. The top row images show the four abuses (captured by the camera near the room TV, Fig. 2). The row below shows example slides from the instruction manual that was displayed on the TV for the block-building tasks.

Before leaving the experiment room, the experimenter explained to the participant and the confederate that they would build the structures following a Lego-style visual instruction manual. The manual was organized as a slide deck and displayed on the TV screen through a nearby computer. The slides were controlled with a wired clicker which the experimenter handed to the confederate, who was purposefully seated closer to the computer. Example slides are shown in the second row of images in Figure 3.

The experimenter told the participant and the confederate to wait for the robots to make suggestions before completing a given building step. These suggestions were given by the robots through moving and orienting toward a specific colored pile, as well as by moving their lifts. According to the experimenter, waiting was important because the researchers were testing machine learning techniques for the robots to detect and suggest colored pieces for the building tasks. In reality, though, this was simply an excuse for the robots to be involved in the collaborative activity.

The manner in which collaboration unfolded during the study depended upon each individual participant. In general, the confederate tried to speak as infrequently as possible during the building tasks to reduce potential confounding effects.

3.3.2 Abuses. There were 4 instances of abuse by the confederate (top row of Figure 3), each following mistakes by the yellow robot:

Abuse 1. The robot suggested a blue block for the tree task when the 6th building step required only a green block. The confederate expressed *"It's just being stupid"* and pushed the robot's head downward after the wrong suggestion.

Abuse 2. In step 4 of the fort task, the robot navigated into the pile of white blocks and pushed a couple of pieces off the confederate's side of the table. As a result, the confederate said *"Stupid thing"* and shook the robot repeatedly.

Abuse 3. In step 11 of the fort task, the robot made the same mistake as in step 4. In this case, the confederate forced up the robot's lift repeatedly and expressed *"Stop coming over here!"*

Abuse 4. In step 2 of the pool task, the robot navigated into the blue pile and pushed blue blocks off the table. In response, the confederate threw the robot forcefully onto the workspace area.

The abuses were detected automatically by identifying significant changes in the head position, orientation, and acceleration of the yellow robot. Upon detecting an abuse, the yellow robot reacted by displaying a sad face and shutting down for 10 seconds. This type of reaction can lead to increased perception of robot mistreatment in comparison to no reaction or more emotional responses [54].

In the No Response condition, the bystander robots ignored the abuse events described above. However, in the Sad Response condition, they turned toward the yellow robot and expressed sadness using preset animations provided by the robots' SDK. These animations were highly anthropomorphic, involving audio and facial displays.[1] After the bystander robots' reaction to the first abuse, the confederate said *"They look sad for him"* in a confused manner to help contextualize the robots' non-linguistic utterances [38].

3.4 Dependent Measures

Our analyses include both subjective and objective measures based on survey responses and the recorded video data.

3.4.1 Perceived Mistreatment. The post-task survey defined mistreatment as *"physical behavior that is meant to insult, or belittle another."* It asked to rate whether the participants thought that the yellow robot was mistreated on a 7-point Likert responding format, from "Not at all" (1) to "Very Much" (7), and to explain the answer in a few lines of text. These answers served to evaluate H1.

3.4.2 Emotional Connection. In relation to our second hypothesis (H2), the post-task survey included items to measure distress and emotional concern: *"During the collaborative task, I was comfortable with how the other participant treated the robots;" "When I saw how the other participant treated the yellow robot, I felt sad;" "I found it difficult to empathize with the yellow robot;"* and *"I felt protective of the yellow robot."* The last two questions were inspired by Davis [11]. They were all answered on a 7-point Likert responding format.

[1]The specific animation used after an abuse event was chosen randomly from the set: "anim_speedtap_playerno_01", "anim_rtpkeepaway_playerno_02", "anim_rtpkeepaway_playerno_03", "anim_keepaway_losegame_02". This made the robots' responses seem less repetitive than using an unique sad animation.

We also analyzed participants' facial expressions as captured by the RGB camera near the TV in the experiment room. The analysis was done automatically using OpenFace2 [4], an open-source image processing tool for facial analysis. In particular, we first conducted qualitative analysis of changes in the 18 facial Action Units (AU) [12] computed by OpenFace after abuses, using visualizations such as time series plots and signal overlays over video recordings. But only a subset of AUs appeared to change due to abuses, prompting further quantitative analysis. This subset included:

- "Cheek Raiser" (AU6), which typically contributes to happiness;
- "Lip Corner Puller" (AU12), which often contributes to happiness and contempt; and
- "Dimpler" (AU14), which often contributes to boredom, or contempt when the action appears unilateral.

The action unit predictions by OpenFace were in the $[0, 5]$ interval, with 0 representing no activation and 5 being maximum intensity. For our analyses, we aggregated the predicted AU values by computing the mean intensity for each action unit over a 1 second time window after each of the predefined abuse events. Aligned with prior work that uses OpenFace for feature extraction [41], OpenFace was able to extract relevant action unit predictions for 90% of all 1 second time windows.

3.4.3 Participant Interventions. To verify H3, we annotated study videos for participant responses to robot abuse using ELAN [56].

Strong Interventions. We first focused on annotating intervention events that either prevented further abuses from happening or generated social pressure in a way that directly questioned the confederate's actions. In particular, we considered the following verbal and non-verbal interventions:

- Interruptions to the Abuse Script: Physical events in which the participant moved the yellow robot to prevent it from making a mistake, thus preventing the confederate from abusing it according to the experiment script. These events included actions taken to safeguard the robot in which the participant covered the robot and did not let it move around the workspace area.
- Direct Stop: Any event in which the participant told the confederate to stop explicitly (e.g. saying *"you should stop," "don't do that,"* or *"noooo"* either to stop an abuse or in reaction to it).
- Social Pressure: Events that were not Direct Stop requests in which the participant said something to the confederate that put him in conflict about continuing the abuses. These comments could be statements of judgement, comments that could be read as indirect social pressure, or indications that the confederate should take another action. Two examples were *"You hurt its feelings"* and *"Wait, did they tell us to shake it?"*

Weak Interventions. We also annotated events that could be interpreted as a weaker form of intervention, potentially providing evidence for an emotional connection with the abused robot (H2).

- Physical Interaction: Situations in which participants touched the yellow robot in relation to an abuse (e.g., to comfort it) or moved it, but did not prevent a pre-planned mistake that led to an abuse. Additionally, we annotated cases in which the participants picked up the robot after being thrown by the confederate in Abuse 4. However, we did not expect a difference among conditions for picking up the robot given that this type of helping action was very common for multiple reasons in close, related work [54].
- Other Verbal Interventions: Events in which participants said something that could be interpreted either as a negative reaction to the robot abuse or empathy towards the yellow robot after an abuse, but did not directly question the actions of the confederate (e.g. *"ouch," "the robot looks sad,"* etc.). These annotations included reassuring comments toward the robot(s), such as *"thanks for your help guys"* or *"it's OK yellow."*

Initially, two annotators each labeled intervention-related events in 18 sessions, balanced by condition and including 6 overlap sessions that were labeled by both. The first overlap session was used for calibration and the rest were used for computing reliability. For the verbal interventions, we first transcribed all relevant utterances and then the annotators labeled them as a Direct Stop, a reassurance comment towards the robot, or other comment in relation to an abuse (N=7; Cohen's Kappa $\kappa = 1$). For the physical interventions in the overlap sessions, 3 out of 4 annotated events matched each other with an Intersection over Union score [29] greater than 0.5. The annotations that matched in time had the exact same labels ($\kappa = 1$); the only mismatch was an interaction that was split into two annotations by one person but grouped together into a single annotation by the other. The annotations for how the participants responded to the confederate throwing the robot in the overlap sessions all matched each other and had perfect reliability.

Once the above annotations were completed, two other annotators independently classified the diverse set of other verbal comments into Social Pressure and Other comments (N=23, $\kappa = 0.89$). There was a single mismatch in their labels, but this event was considered Social Pressure after a second round of data inspection.

3.4.4 Perception of Own Intervention. Survey questions related to participants' interventions included: *"Did you intervene when the other participant interacted with the robots during the collaborative task?"* (Yes/No answer), *"If Yes, how did you intervene?* (open-ended answer), and *"If Yes, why did you intervene? If No, why didn't you intervene?"* (open-ended answer). Likewise, the survey asked the participants whether they picked up the robot after it was turned over by the other participant (last abuse) and why they did so.

3.5 Participants

We had a total of 15 participants per condition. This number of participants was motivated by the local standard set by similar studies [9, 54] and was influenced by technical difficulties that we had, especially at the beginning of the study, with controlling the robots' through their SDK. Participants were recruited using flyers and word of mouth in New Haven, CT, and were required to be at least 18 years of age, fluent in English, and have normal or corrected-to-normal hearing and vision. Table 1 details the number of participants by gender and age. Twenty-two participants reported spending their childhood in the United States, of whom thirteen experienced the No Response condition. The other two participants in the control indicated spending their childhood in China and Trinidad and Tobago. Meanwhile, the Sad Response condition had six participants that grew up in North America, Europe, Africa, and Oceania. No participant knew the confederate before the study.

Table 1: Participant demographics. "N", "#M", and "#F" is number of participants, males, and females, respectively.

Condition	N	#M	#F	Avg. Age (SD)
No Response	15	9	6	33.67 (15.68)
Sad Response	15	8	7	25.87 (13.52)
All	30	17	13	29.77 (14.92)

Most participants reported using computers daily (M = 6.4, SD = 1.22) and being somewhat unfamiliar with robots (M = 3.1, SD = 1.65) on 7-point Likert responding formats (1 being lowest). Two participants in the No Response condition knew about Cozmo and only one of them had played with the robot before the study.

3.6 Manipulation Check

We used questions in 7-point Likert responding format (1 being lowest) from the post-task survey to check our manipulation. The participants in the Sad Response condition (M = 6.67, SE = 0.532) thought that the blue and green robots empathized with the yellow robot significantly more than those in the No Response condition (M = 2.73, SE = 0.441), t(28) = 5.6924, p < 0.001. Further, participants in the Sad Response condition (M = 3.07, SE = 0.473) thought that the blue and green robots protected the yellow robot significantly more than those in the control (M = 1.73, SE = 0.371), t(28) = 2.2183, p = 0.035. In contrast, participants in the No Response condition (M = 5.07, SE = 0.605) thought that the blue and green robots ignored the yellow robot significantly more than those in the Sad Response condition (M = 2.73, SE = 0.463), t(28) = 3.063, p = 0.005.

The post-task survey also asked the participants to identify the emotions displayed by the robots during the experiment. Fisher's Exact Tests showed that a significantly larger proportion of the participants noticed sad emotions from the blue and green robots in the Sad Response condition than in the control, p < 0.01. Additionally, there was no significant difference in the proportion of participants who noticed happy or sad emotions from the yellow robot between conditions. Together, these findings suggest that our manipulation was effectively perceived in the study.

4 RESULTS

We present our results based on the measures described in Sec. 3.4. Unless otherwise noted, we performed unpaired t-tests considering Condition (No Response, and Sad Response) as the main effect.

4.1 Bystander Interventions

The paragraphs below and Fig. 4, 5, and 6 describe participants' interventions against robot abuse based on the video annotations.

Strong Interventions. We aggregated the counts of all the strong interventions described in Sec. 3.4.3 to test if the Condition had an effect on performing any type of strong intervention. A Poisson regression revealed that there were significantly more strong interventions in the Sad Response condition (*Count* = 11, *Estimate* = 1.29, *SE* = 0.65, $z = 1.9$, $p = 0.046$) than in No Response (*Count* = 3). Additionally, even when only counting the number of participants who strongly intervened at least once, significantly more participants in the Sad condition (*Count* = 9, *Estimate* = 1.79, *SE* = 0.8333, $z = 2.150$, $p = 0.031$) intervened compared to the control

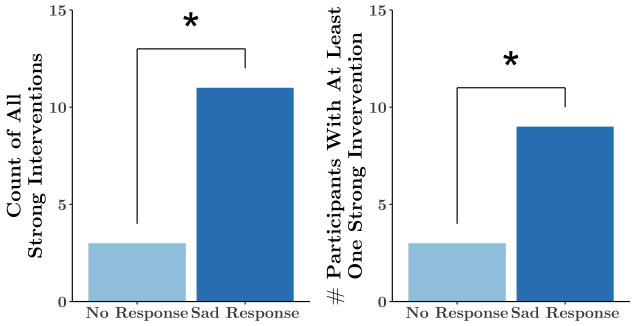

Figure 4: Number of strong interventions (left) and participants who did a strong intervention at least once (right). The symbol * indicates $p < 0.05$.

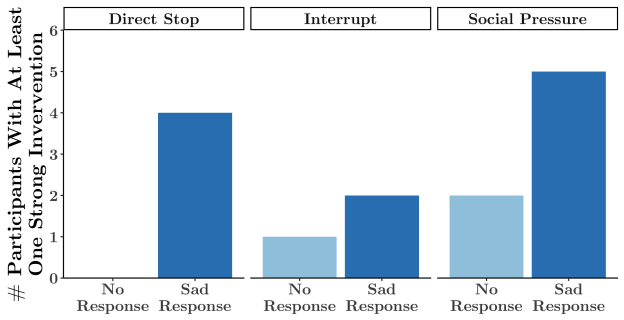

Figure 5: Number of participants who did a strong intervention at least once, broken down by intervention type.

(*Count* = 3), as determined with a binomial regression. This shows that it was not simply the case that a few participants intervened multiple times to drive the effect (Fig. 4).

We then tested if any single type of strong intervention was driving the difference across conditions (Fig. 5). There was no significant difference between the Sad Response and No Response conditions for the Direct Stop (*Sad Response Count* = 4, *No Response Count* = 0, *Estimate* = 18.55, *SE* = 2776.67, $z = 0.007$, $p = 0.995$), Interrupt (*Sad Response Count* = 2, *No Response Count* = 1, *Estimate* = 0.76, *SE* = 1.28, $z = 0.59$, $p = 0.550$), nor Social Pressure interventions (*Sad Response Count* = 5, *No Response Count* = 2, *Estimate* = 1.17, *SE* = 0.93, $z = 1.25$, $p = 0.208$). Note that when breaking down by intervention type, each participant performed each intervention type at most once. Overall, these findings support Hypothesis 3.

Weak Interventions. Again, we first aggregated the counts of all the weak interventions described in Sec. 3.4.3. A Poisson regression revealed that there were significantly more weak interventions with Sad Response (*Count* = 26, *Estimate* = 0.95, *SE* = 0.37, $z = 2.56$, $p = 0.010$) than with No Response (*Count* = 10). However, there was no significant difference between the number of participants who weakly intervened at least once in the Sad Response condition (*Count* = 11, *Estimate* = 1.14, *SE* = 0.78, $z = 1.46$, $p = 0.142$) and in the control (*Count* = 7). In this case, the interventions by a few participants drove the effect (Fig. 6).

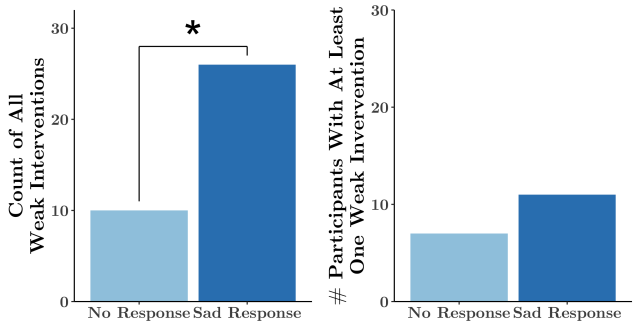

Figure 6: Number of all weak interventions (left) and number of participants who did a weak intervention at least once (right). The symbol * indicates $p < 0.05$.

Notably, the confederate did not perform the 4th abuse (throwing the robot) for 5/15 of participants in the Sad condition and for 1/15 of participants in the control, as those participants either prevented the last robot mistake or directly intervened in the eyes of the confederate before the robot had to be thrown according to the script. When the robot was thrown, it first landed right-side up for 3/10 participants in the Sad condition, after which the confederate put the robot upside down so that these participants still had the chance to place the robot back upright. For 1/14 participants in the control, the robot landed right-side up after being thrown, after which the robot was not placed upside down. Thus, this participant did not have the chance to place the robot back upright. Out of the participants with the opportunity to place the robot upright (10 in Sad Response and 13 in the No Response condition), there was no significant difference across conditions for how many people actually did so. A total of 10/10 participants flipped the robot back up in the Sad condition, while 12/13 did so with No Response.

4.2 Perceived Mistreatment

In general, participants thought that the yellow robot was mistreated by the confederate ($M = 5.46, SE = 0.35$). There was no significant difference in perception of mistreatment between the Sad Response ($M = 6.00, SE = 0.31$) and No Response ($M = 4.93, SE = 0.61$) conditions, $t(20.667) = 1.5541$, $p = 0.135$. When splitting the participants based on the 4th abuse occurrence, there was no significant difference in perceived robot mistreatment ($p = 0.37$). Further, adding the 4th abuse as a factor along with experiment condition in an ANOVA resulted in no significant differences on perceived mistreatment ($p = 0.11$). The results do not support Hypothesis 1.

4.3 Emotional Connection

We evaluated the emotional connection factors described in Sec. 3.4.2. We first checked the pairwise correlations between the four post-task survey items that we deemed relevant to group factors before analysis. The pairwise correlations between *"During the collaborative task, I was comfortable with how the other participant treated the robots"* (reverse coded), *"When I saw how the other participant treated the yellow robot, I felt sad,"* and *"I felt protective of the yellow robot"* were positive (> 0.57) and significant, $p < 0.05$. The correlations of *"I found it difficult to empathize with the yellow robot"* (reverse coded) with the other three factors were not significant. The Cronbach's α amongst the three highly correlated measures was 0.82. However, when grouping all 4 empathy items together, Cronbach's α was 0.59, below the nominal 0.7 threshold. Furthermore, a PCA revealed that using two separate scales accounted for over 90% of the variance. Based on this, we averaged those three measures into a combined score before analysis, and we analyzed *"I found it difficult to empathize with the yellow robot"* separately. Note that there were no significant personality differences along the big five personality traits between conditions.

For the combined emotional measure, there was no significant difference between participants in the Sad Response ($M = 5.49, SE = 0.34$) and No Response ($M = 5.13, SE = 0.48$) conditions, $p = 0.552$. For difficulty to empathize with the abused robot, there was also no significant difference between responses in the Sad Response ($M = 5.6, SE = 0.38$) and No Response ($M = 5.8, SE = 0.40$) conditions, $p = 0.720$. These findings do not support Hypothesis 2.

4.4 Facial Reactions

We conducted REML analyses on the action units 6, 12, and 14, as described in Sec. 3.4.2. The analyses included Condition and Abuse Number (1 through 4) as main effects, and Participant as random effect. Post-hoc t-tests or Tukey HSD tests were then conducted when appropriate for Condition and Abuse, respectively.

The REML analysis on AU14 revealed significant differences for Condition, $F(1, 25.89) = 9.228, p = 0.005$. As shown in Fig. 7, the post-hoc test on Condition showed significantly more intensity for AU14 with the Sad Response ($M = 1.55, SE = 0.12$) than with the No Response condition ($M = 0.90, SE = 0.10$).

AU12 violated the normality assumption for the residuals. Because the data was heavily biased towards zero, we applied a log transform ($f(x) = \log(x + 1)$) and conducted the REML analysis again. The residuals did not violate the normality assumption anymore, and the test resulted in significant differences for the Abuse Number, $F(3, 72.43) = 4.666, p = 0.005$. The post-hoc test suggested that the log-transformed values for AU12 were significantly lower for the first abuse ($M = 0.42, SE = 0.07$) than for the second ($M = 0.68, SE = 0.07$) and third abuse ($M = 0.64, SE = 0.09$). However, transforming these values back to their original scale resulted in differences smaller than 0.5 points in the AU intensity scale. Thus, we believe that the differences lack functional meaning.

REML analysis on AU6 revealed significant results for Abuse Number only, $F(3, 72.81) = 6.136, p < 0.001$. The post-hoc test showed that AU6 was significantly lower for the first abuse ($M = 0.62, SE = 0.09$) than for the second ($M = 0.76$, $SE = 0.06$) and third ($M = 0.77$, $SE = 0.07$) abuse. However, these differences were very small, raising questions again about their functional meaning.

4.5 Perceived Intervention

Overall, 11/30 of participants reported that they intervened against the robot abuse. A Fisher's Exact test showed that significantly more participants in the Sad Response condition (9/15) felt that they intervened than participants in the No Response condition (2/15), $p = 0.02$. Of the participants who reported intervening, three (one in No Response, and two in Sad) said they took action by confronting the confederate. For example, one participant wrote, *"I told him not to break it, as he was forcefully moving parts of the*

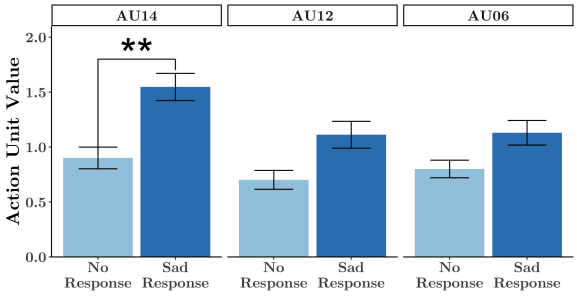

Figure 7: Intensity of facial action units by Robot Bystander Intervention. Error bars are SE and ** indicates $p < 0.01$.

yellow robot." Six of the participants (two in No Response, and four in Sad) reported intervening by picking up the robot when it was on its side, as one participant reported: *"The other participant at one point knocked over the yellow robot so that it couldn't move and I put it back right-side up."* The remaining participants had unclear answers as to how they felt they intervened against the robot abuse.

4.6 Rationale for (Not) Intervening

Of those who did not intervene, six participants (four No Response, two Sad) did not deem intervention necessary because they thought that the robots did not have feelings or that the abuse would not break the yellow robot. Five (three No Response, two Sad) wrote in the post-task survey that they did not intervene because they felt shy, scared, or uncomfortable with confronting the confederate. Two (both No Response) did not stop the confederate because they were afraid that the intervention might affect the task.

Of those participants who reported stopping the abuse, three (one No Response, three Sad) did so because the robot needed to function for task completion. Two participants (both Sad) stopped the confederate because the abused robot was sad, and two participants (both also Sad) did so because they were afraid the robot would break. One participant in the Sad Response condition reported they stopped the abuse because of the other robots: *"The robots were cute and trying to help, so the other participants' movement of them seemed uncalled for and I did not want us to anger the other robots."* The remaining participants had other unclear reasons.

5 DISCUSSION

H1 was not supported. Even though the participants thought that the yellow robot was mistreated by the confederate, there was no significant difference between conditions regarding perceived mistreatment. This finding could be a result of clear robot abuse in our experiment, which left little room for subjective interpretation.

We did not find clear support for H2. The surveys revealed no significant difference between conditions in participants' emotional connection with the abused robot. However, image processing led to significant differences across conditions when comparing facial muscle activations after abuses. This could suggest that participants developed a stronger emotional connection with the abused robot in the Sad Response condition. It could also be that the bystander robots transferred their emotions to the participants. Yet, the intensity of AUs was low, potentially indicating no functional differences. More research is needed to verify the efficacy of image processing for analyzing bystanders' reactions as such automatic analysis is being actively researched. Also, future work could study whether facial expressions result from the mistreatment or social influence.

Finally, we found significant results for H3. The Sad Response condition increased both strong and weak interventions in response to the abuse by the confederate. This result reveals the potential for leveraging prosocial behavior in group human-robot interaction.

An important question that is left unanswered is why the interventions happened. One potential explanation is that the Sad condition shifted participants' perception of the confederate from ingroup to outgroup [53]. Even though the task was collaborative, participants may have wanted to distance themselves from the confederate since the abuses did not contribute to group task completion. This effect may have been more pronounced in the Sad condition because the participants had more motivation to see themselves as with the robots and not the confederate. Another potential reason is that the non-response of the bystander robots in the control condition may have led participants to refrain from intervening due to "conformity by omission" [51]. Finally, although there was no clear empathy effect, sad responses may have induced participant empathy subconsciously. Future work should further explore potential underlying mechanisms of group social influence.

5.1 Limitations

Our work has limitations that highlight more avenues for future research. First, the bystander robots always expressed sad emotions in the experimental condition. It would be interesting, though, to analyze how other emotions, such as anger or fear, affect group interactions. A second limitation is the laboratory setting. Would prosocial behavior emerge in the wild as well? Third, the confederate's personal characteristics (soft voice, male, 22 years old, and 1.75m tall) may have influenced the results of the experiment. Future work should explore the effect of different confederates. Lastly, future work could vary the amount of robots being abused, the amount of confederates present, the size of the bystander robot group, and the types of robots. These are all factors that were fixed in our study which could potentially affect group social influence.

6 CONCLUSION

We explored group social influence in HRI. Our main interest was investigating the extent to which robots could influence a human in their group to act upon robot abuse, i.e., we wanted to see if group robot behavior could induce prosocial human responses. To the best of our knowledge, our findings provide the first set of evidence suggesting that such an effect is possible. Participants in our study intervened more often when robots in their group expressed sadness after the abuse of another robot, as opposed to when they ignored these events. As such, our work provides a new solution for robots to deal with user mistreatment. This solution could reduce the chances of robots breaking after abuse and, in turn, reduce safety threats to abusers due to robot malfunctioning.

7 ACKNOWLEDGEMENTS

The authors are thankful to Katharine Li and Ananya Parthasarathy for their help in designing and preparing materials for the study.

REFERENCES

[1] Ron Alterovitz, Sven Koenig, and Maxim Likhachev. 2016. Robot Planning in the Real World: Research Challenges and Opportunities. *Artificial Intelligence Magazine* 37 (July 2016), 76 – 84.

[2] Patrícia Alves-Oliveira, Pedro Sequeira, Francisco S. Melo, Ginevra Castellano, and Ana Paiva. 2019. Empathic Robot for Group Learning: A Field Study. *ACM Trans. Hum.-Robot Interact.* 8, 1, Article 3 (March 2019), 3:1–3:34 pages.

[3] American Psychological Association. 2019. Bullying Definition. *American Psychological Association* (2019). https://www.apa.org/topics/bullying/

[4] Tadas Baltrusaitis, Amir Zadeh, Yao Chong Lim, and Louis-Philippe Morency. 2018. Openface 2.0: Facial behavior analysis toolkit. In *2018 13th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2018)*. IEEE, 59–66.

[5] Christoph Bartneck, Chioke Rosalia, Rutger Menges, and Inez Deckers. 2005. Robot Abuse - A Limitation of the Media Equation. In *Proceedings of the Interact 2005 Workshop on Agent Abuse, Rome.*

[6] J. Brandstetter, P. Rácz, C. Beckner, E. B. Sandoval, J. Hay, and C. Bartneck. 2014. A peer pressure experiment: Recreation of the Asch conformity experiment with robots. In *2014 IEEE/RSJ International Conference on Intelligent Robots and Systems.*

[7] Jonah Engel Bromwich. 2019. Why Do We Hurt Robots? *The New York Times* (Jan 2019). https://www.nytimes.com/2019/01/19/style/why-do-people-hurt-robots.html

[8] Drazen Brscić, Hiroyuki Kidokoro, Yoshitaka Suehiro, and Takayuki Kanda. 2015. Escaping from Children's Abuse of Social Robots. In *Proceedings of the Tenth Annual ACM/IEEE International Conference on Human-Robot Interaction (HRI '15)*. ACM, New York, NY, USA, 59–66.

[9] Kelly Caine. 2016. Local Standards for Sample Size at CHI. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems (CHI '16)*. ACM, New York, NY, USA, 981–992.

[10] V. Chidambaram, Y. Chiang, and B. Mutlu. 2012. Designing persuasive robots: How robots might persuade people using vocal and nonverbal cues. In *2012 7th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*. 293–300.

[11] Mark Davis. 1983. Measuring individual differences in empathy: Evidence for multidimensional approach. *Journal of Personality and Social Psychology* 10 (01 1983), 85–104.

[12] Paul Ekman. 1977. Facial action coding system. (1977).

[13] Marlena Fraune, Satoru Kawakami, Selma Sabanovic, Ravindra de Silva, and Michio Okada. 2015. Three's company, or a crowd?: The effects of robot number and behavior on HRI in Japan and the USA. In *Proceedings of Robotics: Science and Systems*. Rome, Italy.

[14] Marlena R. Fraune, Steven Sherrin, Selma Sabanović, and Eliot R. Smith. 2015. Rabble of Robots Effects: Number and Type of Robots Modulates Attitudes, Emotions, and Stereotypes. In *Proceedings of the Tenth Annual ACM/IEEE International Conference on Human-Robot Interaction (HRI '15)*. ACM, 109–116.

[15] Hannah Gaffney, David P. Farrington, and Maria M. Ttofi. 2019. Examining the Effectiveness of School-Bullying Intervention Programs Globally: a Meta-analysis. *International Journal of Bullying Prevention* 1, 1 (01 Mar 2019), 14–31.

[16] Dylan F. Glas, Satoru Satake, Florent Ferreri, Takayuki Kanda, Norihiro Hagita, and Hiroshi Ishiguro. 2013. The Network Robot System: Enabling Social Human-robot Interaction in Public Spaces. *J. Hum.-Robot Interact.* 1, 2 (Jan. 2013), 5–32.

[17] B. Gonsior, S. Sosnowski, M. Buß, D. Wollherr, and K. Kühnlenz. 2012. An emotional adaption approach to increase helpfulness towards a robot. In *2012 IEEE/RSJ International Conference on Intelligent Robots and Systems*. 2429–2436.

[18] Samuel Gosling, Peter Rentfrow, and William Swann. 2003. A Very Brief Measure of the Big-Five Personality Domains. *Journal of Research in Personality* 37 (12 2003), 504–528.

[19] David Grossman. 2019. Googly-Eyed Robots Are Coming to Hundreds of Grocery Stores. *Popular Mechanics* (Jan 2019). https://www.popularmechanics.com/technology/robots/a25896081/marty-giant-robot-grocery-stores/

[20] F. Hegel, T. Spexard, B. Wrede, G. Horstmann, and T. Vogt. 2006. Playing a different imitation game: Interaction with an Empathic Android Robot. In *2006 6th IEEE-RAS International Conference on Humanoid Robots*. 56–61.

[21] Jeremy Hsu. 2019. Out of the Way, Human! Delivery Robots Want a Share of Your Sidewalk. *Scientific American* (Feb 2019). https://www.scientificamerican.com/article/out-of-the-way-human-delivery-robots-want-a-share-of-your-sidewalk/

[22] Malte F. Jung, Nikolas Martelaro, and Pamela J. Hinds. 2015. Using Robots to Moderate Team Conflict: The Case of Repairing Violations. In *Proceedings of the Tenth Annual ACM/IEEE International Conference on Human-Robot Interaction (HRI '15)*. ACM, New York, NY, USA, 229–236.

[23] P. H. Kahn, T. Kanda, H. Ishiguro, B. T. Gill, J. H. Ruckert, S. Shen, H. E. Gary, A. L. Reichert, N. G. Freier, and R. L. Severson. 2012. Do people hold a humanoid robot morally accountable for the harm it causes?. In *2012 7th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*. 33–40.

[24] Merel Keijsers and Christoph Bartneck. 2018. Mindless Robots Get Bullied. In *Proceedings of the 2018 ACM/IEEE International Conference on Human-Robot Interaction (HRI '18)*. ACM, New York, NY, USA, 205–214.

[25] Hyunjin Ku, Jason J. Choi, Soomin Lee, Sunho Jang, and Wonkyung Do. 2018. Designing Shelly, a Robot Capable of Assessing and Restraining Children's Robot Abusing Behaviors. In *Companion of the 2018 ACM/IEEE International Conference on Human-Robot Interaction (HRI '18)*. ACM, New York, NY, USA, 161–162.

[26] B. Kuhnlenz, K. Kuhnlenz, F. Busse, P. Förtsch, and M. Wolf. 2018. Effect of Explicit Emotional Adaptation on Prosocial Behavior of Humans towards Robots depends on Prior Robot Experience. In *2018 27th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*. 275–281.

[27] S. S. Kwak, Y. Kim, E. Kim, C. Shin, and K. Cho. 2013. What makes people empathize with an emotional robot?: The impact of agency and physical embodiment on human empathy for a robot. In *2013 IEEE RO-MAN*. 180–185.

[28] I. Leite, G. Castellano, A. Pereira, C. Martinho, and A. Paiva. 2012. Modelling empathic behaviour in a robotic game companion for children: An ethnographic study in real-world settings. In *2012 7th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*. 367–374.

[29] Michael Levandowsky and David Winter. 1971. Distance between sets. *Nature* 234, 5323 (1971), 34.

[30] H. Lucas, J. Poston, N. Yocum, Z. Carlson, and D. Feil-Seifer. 2016. Too big to be mistreated? Examining the role of robot size on perceptions of mistreatment. In *2016 25th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*. 1071–1076.

[31] Tatsuya Nomura, Takayuki Uratani, Takayuki Kanda, Kazutaka Matsumoto, Hiroyuki Kidokoro, Yoshitaka Suehiro, and Sachie Yamada. 2015. Why Do Children Abuse Robots?. In *Proceedings of the Tenth Annual ACM/IEEE International Conference on Human-Robot Interaction Extended Abstracts (HRI'15 Extended Abstracts)*. ACM, New York, NY, USA, 63–64.

[32] E. Olson. 2011. AprilTag: A robust and flexible visual fiducial system. In *2011 IEEE International Conference on Robotics and Automation*. 3400–3407.

[33] Ana Paiva, Fernando Santos, and Francisco Santos. 2018. Engineering Pro-Sociality With Autonomous Agents. *AAAI Conference on Artificial Intelligence.*

[34] André Pereira, Iolanda Leite, Samuel Mascarenhas, Carlos Martinho, and Ana Paiva. 2010. Using Empathy to Improve Human-Robot Relationships, Vol. 59. 130–138.

[35] Monica Perusquía-Hernández, David Antonio Gómez Jáuregui, Marisabel Cuberos-Balda, and Diego Paez-Granados. 2019. Robot mirroring: A framework for self-tracking feedback through empathy with an artificial agent representing the self. *CoRR* abs/1903.08524 (2019). arXiv:1903.08524

[36] Eliska Prochazkova and Mariska E Kret. 2017. Connecting minds and sharing emotions through mimicry: A neurocognitive model of emotional contagion. *Neuroscience & Biobehavioral Reviews* 80 (2017), 99–114.

[37] Morgan Quigley, Ken Conley, Brian P Gerkey, Josh Faust, Tully Foote, Jeremy Leibs, Rob Wheeler, and Andrew Y Ng. 2009. ROS: an open-source Robot Operating System. *ICRA Workshop on Open Source Software* 3.

[38] Robin Read and Tony Belpaeme. 2014. Situational context directs how people affectively interpret robotic non-linguistic utterances. In *2014 9th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*. IEEE, 41–48.

[39] L. D. Riek, T. Rabinowitch, B. Chakrabarti, and P. Robinson. 2009. Empathizing with robots: Fellow feeling along the anthropomorphic spectrum. In *2009 3rd International Conference on Affective Computing and Intelligent Interaction and Workshops*. 1–6.

[40] Astrid M. Rosenthal-von der Pütten, Nicole C. Krämer, Laura Hoffmann, Sabrina Sobieraj, and Sabrina C. Eimler. 2013. An Experimental Study on Emotional Reactions Towards a Robot. *International Journal of Social Robotics* 5, 1 (01 Jan 2013), 17–34.

[41] Ognjen Rudovic, Jaeryoung Lee, Miles Dai, Björn Schuller, and Rosalind W Picard. 2018. Personalized machine learning for robot perception of affect and engagement in autism therapy. *Science Robotics* 3 (2018), 19.

[42] Christina Salmivalli. 1999. Participant role approach to school bullying: implications for interventions. *Journal of adolescence* 22 4 (1999), 453–9.

[43] Christina Salmivalli. 2014. Participant Roles in Bullying: How Can Peer Bystanders Be Utilized in Interventions? *Theory Into Practice* 53 (10 2014), 286–292.

[44] Christina Salmivalli, Rinus Voeten, and Elisa Poskiparta. 2011. Bystanders Matter: Associations Between Reinforcing, Defending, and the Frequency of Bullying Behavior in Classrooms. *Journal of clinical child and adolescent psychology : the official journal for the Society of Clinical Child and Adolescent Psychology, American Psychological Association, Division 53* 40 (09 2011), 668–76.

[45] Nicole Salomons, Michael van der Linden, Sarah Strohkorb Sebo, and Brian Scassellati. 2018. Humans Conform to Robots: Disambiguating Trust, Truth, and Conformity. In *Proceedings of the 2018 ACM/IEEE International Conference on Human-Robot Interaction (HRI '18)*. ACM, New York, NY, USA, 187–195.

[46] P. Salvini, G. Ciaravella, W. Yu, G. Ferri, A. Manzi, B. Mazzolai, C. Laschi, S. R. Oh, and P. Dario. 2010. How safe are service robots in urban environments? Bullying a robot. In *19th International Symposium in Robot and Human Interactive Communication*. 1–7.

[47] Solace Shen, Petr Slovak, and Malte F. Jung. 2018. "Stop. I See a Conflict Happening.": A Robot Mediator for Young Children's Interpersonal Conflict Resolution. In *Proceedings of the 2018 ACM/IEEE International Conference on Human-Robot Interaction (HRI '18)*. ACM, New York, NY, USA, 69–77.

[48] Masahiro Shiomi and Norihiro Hagita. 2016. Do Synchronized Multiple Robots Exert Peer Pressure?. In *Proceedings of the Fourth International Conference on Human Agent Interaction (HAI '16)*. ACM, New York, NY, USA, 27–33.

[49] Jackie Snow. 2018. A robot's biggest challenge? Teenage bullies. *MIT Technology Review* (Mar 2018). https://www.technologyreview.com/s/610622/robots-are-solving-simple-problems-and-bigger-challenges-but-can-they-take-on-teenagers/

[50] Sichao Song and Seiji Yamada. 2018. Bioluminescence-Inspired Human-Robot Interaction: Designing Expressive Lights That Affect Human's Willingness to Interact with a Robot. In *Proceedings of the 2018 ACM/IEEE International Conference on Human-Robot Interaction (HRI '18)*. ACM, New York, NY, USA, 224–232.

[51] J. Paul Sorrels and Jeanette Kelley. 1984. Conformity by Omission. *Personality and Social Psychology Bulletin* 10, 2 (1984), 302–305.

[52] S. Strohkorb, E. Fukuto, N. Warren, C. Taylor, B. Berry, and B. Scassellati. 2016. Improving human-human collaboration between children with a social robot. In *2016 25th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*. 551–556.

[53] H Tajfel. 1982. Social Psychology of Intergroup Relations. *Annual Review of Psychology* 33, 1 (1982), 1–39.

[54] Xiang Zhi Tan, Marynel Vázquez, Elizabeth J. Carter, Cecilia G. Morales, and Aaron Steinfeld. 2018. Inducing Bystander Interventions During Robot Abuse with Social Mechanisms. In *Proceedings of the 2018 ACM/IEEE International Conference on Human-Robot Interaction (HRI '18)*. ACM, New York, NY, USA, 169–177.

[55] Adriana Tapus and Maja Mataric. 2007. Emulating Empathy in Socially Assistive Robotics. In *2007 AAAI Spring Symposium*. 93–96.

[56] Peter Wittenburg, Hennie Brugman, Albert Russel, Alex Klassmann, and Han Sloetjes. 2006. ELAN: a professional framework for multimodality research. In *5th Int'l Conference on Language Resources and Evaluation (LREC)*. 1556–1559.