Who's Laughing NAO? Examining Perceptions of Failure in a Humorous Robot Partner

Haley N. Green[1], Md Mofijul Islam[1], Shahira Ali[2], Tariq Iqbal[1]
[1]School of Engineering and Applied Science, University of Virginia, Charlottesville, USA
[2]Department of Cognitive Science, University of Virginia, Charlottesville, USA
{hng9vf, mi8uu, sma6zw, tiqbal}@virginia.edu

***Abstract*—Social robots are being deployed to interact with people in various scenarios, where they are expected to incorporate human-like conversational strategies to achieve fluency in interactions. For example, current robots are designed to perform advanced communication strategies (i.e., personal anecdotes, explanations, and apologies) to recover from task failure. However, these tactics are not always sufficient for failure recovery as they can be lengthy and insufficient for encouraging future interactions. In human-human interactions, people often use humor as a low-risk and engaging method for managing failures. Thus, the successful execution of advanced, human-like humor could enable robots to recover from task failures more efficiently. In this paper, we present a human-robot interaction study exploring how a robot's utilization of various human-like humor types (i.e., affiliative, aggressive, self-enhancing, and self-defeating) are perceived by human teammate ($n = 32$) and an external observer of the interaction ($n = 256$). Additionally, we have explored the effects of performance, humor type, perspective, and previous experience with robots on the participants' perceptions of warmth, competence, and the robot as a teammate. Our results indicate that dyadic participants rated the successful robot to be more *competent* and a better *teammate* than the bystander participants. Additionally, the results indicate that participants with less experience with robots found the successful robot to be more *competent* than participants with high levels of experience. These findings will enable the human-robot interaction community to develop more engaging robots for fluent interactive experiences in the future.**

***Index Terms*—humor; failure; recovery; human-robot interaction**

I. Introduction

Recent advancements in the capabilities of social robots have enabled them to be increasingly useful as collaborative partners. Specifically, robots are able to engage with humans in various environments, such as healthcare, customer service, and education [1]–[9]. The increase in day-to-day human-robot interactions has led to a surplus in efforts to improve the interactive capabilities of robots [10]–[18]. In particular, robots are being designed to employ more human-like communication strategies to facilitate interactions with human agents. One of the popular domains for communication lies in failure recovery [19]–[21]. For example, robots are using techniques, such as expressive emotions, explainability, and personal anecdotes to recover from failures in an interaction [22]–[27]. However, when failure occurs, these strategies can be lengthy. Consequently, there is a need for an engaging recovery tactic that encourages further interactions and communication.

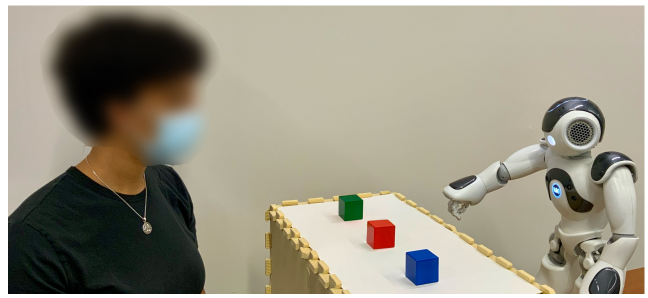

Fig. 1. A NAO robot plays the *iSpy* game with a human. The participant sits opposite the NAO during the game and instructs the robot to identify the green, red, or blue block. Here, the robot correctly identifies the green block during the game.

In human-human interactions, humans have been known to utilize humor as a failure recovery tool [28], [29]. Specifically, humans often employ different *types* of humor when attempting to mitigate failure and conflict. People often incorporate their past experiences, task awareness, and context to generate more situationally relevant jokes. In addition to defusing conflict and mitigating failure, this more advanced humor strategy can also demonstrate the human's personality and task-related competence. However, the efficacy of humor for human-robot interaction remains an open research question. In previous works, the joke-telling capabilities of social robots consist of puns and "knock-knock" jokes [30], [31]. While these forms of humor have been used to establish the robot as a more human-like and sociable interaction partner, they are not always task-related and do not necessarily foster collaborative bonds or bolster the robot's capabilities. To make robots more fluent interactive partners, it is essential to perform systematic investigations to observe how robots' utilization of various types of humor is perceived as a failure recovery strategy.

When developing failure recovery strategies, it is also important to note that human-robot interactions are not always isolated occurrences [12], [32]–[34]. While a robot may be designed to engage in dyadic interactions, there is the potential for bystanders to observe the interaction and formulate their own perceptions about the robot. Therefore, a robot's failure and subsequent recovery tactic could influence more than just the direct interaction partner. As a result, it is important to consider how bystanders may perceive robots and whether those perceptions vary from agents that are directly engaging with the robot. To our knowledge there has not been a

978-1-6654-0731-1/22/$31.00 ©2022 IEEE

systematic study on how perspective affects perceptions of a robot employing humor in a human-robot interaction scenario.

In this work, we are interested in exploring whether a robot's use of more advanced, human-like humor can mitigate task failure. Specifically, we aim to examine whether a humorous failing robot would be considered as warm, as competent, and as good of a teammate as a robot that does not experience failure and a robot that experiences failure but does not employ humor as a mitigation strategy. Additionally, we examine whether factors, such as observer perspective (*dyadic* vs. *bystander*) and previous robot experience affect how humor is perceived. We have designed a human-robot interaction scenario, modeled after the children's game, *iSpy*, where a NAO robot performs a joke in response to a mistake made during a block identification task. To engender a more fluent and human-like interaction, we designed the robot to incorporate a specific humor type (i.e., *affiliative*, *aggressive*, *self-enhancing* and *self-defeating*). We examine the effects of overall humor (humor vs. no-humor), humor type (four different humor types), perspective (dyadic vs. bystander), performance (successful vs. fail with and without humor), and previous robot experience (no to expert-level experience) on ratings of *warmth*, *competence*, and the robot as a *teammate*. We performed a between-subjects study in which we collected dyadic ($n = 32$) and bystander ($n = 256$) perceptions of the robot's failure mitigation approach.

The results suggest that bystanders found the successful robot and robot that failed and used affiliative humor to be more *competent* than the robots that failed and used no humor, whereas the dyadic participants only found the robot that never experienced failure to be significantly more *competent* than the robots that failed and used no humor. Moreover, the results suggest that dyadic participants rated the successful robot to be more *competent* and a better *teammate* than the bystander participants. We did not observe a significant effect of humor and previous robot experience on perceptions of warmth, competence and the robot as a teammate. However, the results indicate a significant effect of previous robot experience and performance on perceptions of the robot's *competence*. Generally, participants with less experience with robots found the successful robot to be more *competent* than participants with high levels of experience. These findings will allow us to develop more engaging robots and collaborative human-robot experiences in the future.

II. Related Work

A. Humor in Human-Human Interactions

Humor has been examined in human-human interactions (HHI) as a conflict resolution and stress-relief strategy [35]–[39]. For example, Kobel and Groeppel-Klein found that humor could be employed as a method for reducing frustration and tension in customer service [28]. Moreover, Mesmer-Magnus et al. observed the different health-related benefits of positive humor in the workplace [40]. Previous work has also explored varying factors that can affect the perception of humor in HHI. For example, Norrick and Spitz observed that conflict severity and relationship power dynamics affect the perception of humor in HHI [29]. Additionally, Bippus explored the various internal and external motives behind humor usage in a conflict management [41]. Finally, humor has also been used to foster connections, particularly in office environments. For instance, Warren, Barsky, and McGraw examined whether humor could be used to achieve consumer goals while enhancing working relationships [42].

B. Humorous Robots

There has also been some previous research on incorporating humor into social robots. In particular, the timing and adaptivity of a humorous robot has been explored in the context of stand-up comedy [43]–[45]. Moreover, Nijholt also surveyed a social humanoid robot in the context of stand-up comedy to explore the effects of timing and adaptivity on audience reception [46]. Additionally, Weber et al. examined how a robot can analyze reactions to humor and adapt its jokes to better align with the human's sense of humor [47]. In HRI, humor has also been used as a conflict resolution strategy. For example, Stoll et al. examined perceptions of humor type in a robot versus human conflict mediator [48].

Recently, social robots have utilized humor to achieve certain engagement goals [49], [50]. For example, Oliveira et al. highlighted the various physical and psychological benefits of humor when used to mitigate the effect of failure [51]. Furthermore, Sebo et al. explored whether a robot could use humor to encourage expressions of trust-related behavior [52]. Additionally, Adamson et al. leveraged humor in a robot-photographer to elicit spontaneous smiles [53].

C. Failure Mitigation Strategies in Social Robots

In any interaction, there are many unforeseen external or environmental variables that can lead to different types of failure [54]. Consequently, in real-world settings, it is impossible to anticipate the nuances of every interaction. Even the most capable robots can experience failure, and it is necessary to incorporate mitigation approaches to minimize the negative effects of failures when they occur. In the past, social robots have been designed to employ a variety of failure mitigation tactics that were useful in human-human interactions. Approaches include: explaining errors, apologizing, asking for help, and even preemptively setting partner expectations [19]–[21]. There has also been past research on robot failure and how the various mitigation strategies are perceived [55]–[61].

D. Multi-Human Groups in Human-Robot Interactions

In HRI, another growing field lies in multi-human teams [62]–[66], and there have been past studies on how multiple humans perceive social robots [26], [33], [52], [67]. In particular, robots have been enlisted in groups to address conflict. For example, Jung, Martelaro, and Hinds examined the effects of a robot moderator on perceptions of conflict [34]. Robots have also been used in groups to encourage communication and engagement [68]. Specifically, Trager et al. observed how a social robot can influence intergroup communication and inspire more evenly distributed conversations [19]. Moreover, robots have been used in multi-human groups to encourage competition. For instance, Vázquez et al. examined how a robot could utilize deception to increase players' motivation by reducing the probability of losing [32].

In human-human interactions, humor as a mitigation tactic supplies an attractive alternative to lengthy explanations and unimaginative apologies. When used appropriately, humor can strengthen working relationships even in cases of task failure. In HRI, humor has been previously utilized as a conversational tool to foster bonds and engage users. However, there has been less exploration on how robots are perceived when they employ more nuanced, human-like humor strategies to mitigate failure. Moreover, to our knowledge, there has not been a previous examination on how perspective influences the reception of different humorous strategies. In this work, we aim to examine the effects of more advanced humor, and perspective on perceptions of a robot in a task-failure scenario.

III. Research Questions

We developed a set of research questions to examine perceptions from both a dyadic interaction partner and a bystander.

- **RQ1:** *How do perceptions of warmth, competence, and the robot as a teammate vary for a humorous and non-humorous robot, and how do those perceptions vary with perspective (dyadic vs. bystander)?* In particular, we aim to examine whether perceptions vary for a robot that: (1) successfully completes its task; (2) fails and uses humor as a recovery strategy; or (3) fails and does not use humor to recover. We refer to these three conditions as the robot's *general performance*. Furthermore, we aim to observe whether perspective affects the perceptions of the different failure mitigation scenarios.
- **RQ2:** *What is the effect of humor type on perceptions of warmth, competence and the robot as a teammate, and how do those perceptions vary with perspective?* Specifically, we aim to observe whether perceptions vary for a robot that: (1) successfully completes its task, (2) fails and uses one of the established humor types (affiliative, aggressive, self-defeating, or self-enhancing), or (3) fails and does not use humor. We refer to these three conditions as the robot's *type-specific performance*. Additionally, we aim to observe the interaction of perspective for the more nuanced failure mitigation scenarios.
- **RQ3:** *What is the effect of previous robot experience on perceptions of warmth, competence, and the robot as a teammate in a humorous and non-humorous robot?* Next, we consider whether previous robot experience has an effect on how a robot is perceived when it is successful, fails with humor, and fails without humor.
- **RQ4:** *What is the effect of humor type and previous robot experience on perceptions warmth, competence, and the robot as a teammate?* Finally, we are interested in observing whether previous robot experience affects how the various humor types are received. To examine this, we compare ratings of warmth, competence, and teammate for a robot that successfully completes the task, fails and uses one of the four humor types, or fails and does not use humor.

TABLE I
Sample set of jokes for each of the four humor types.

Humor Type	Joke Example
Affiliative	*Uh oh, I hope they don't dock my pay.*
Aggressive	*Interesting, I guess we're both bad at this game.*
Self-defeating	*This must be why the other robot's don't like to play with me.*
Self-enhancing	*One for two, that's a pass on an engineering curve.*

IV. Method

A. The Four Humor Types

In this study, we utilize the four humor types most commonly observed in human-human interactions [69]. In particular, *affiliative* humor is a style that is affirming to one's self and others. This low-stakes, non-hostile style is performed with the intention of amusing everyone and is typically associated with good-naturedness. Contrarily, *aggressive* humor is often associated with sarcasm and teasing. This style typically involves putting down or ridiculing others. Furthermore, *self-defeating* humor sets the joke-teller as the target of derision in an attempt to amuse others in the process. This style takes "laughing-at-oneself" a step further through excessive self-deprecation. Finally, *self-enhancing* humor, much like affiliative humor, is good-natured. However, this playful, "laugh-at-oneself" style is more introspective and less focused on cultivating relationships. The four humor types encompass the conversational objectives that can arise in an interaction. These objectives include: enhancing the self or the relationships with others by building up or tearing down oneself or others. A sample joke from each humor category is presented in Table I.

B. The Robot

We utilized SoftBank Robotics' NAO robot [70]. The humanoid robot is approximately 58cm in height. In this study, we combined the expressive behavior modules in SoftBank's Choregraphe suite with Amazon Polly's text-to-speech platform [71]. We elected to use Amazon Polly over Choregraphe's text-to-speech option because of Amazon Polly's robust annunciation and timing capabilities [43]–[45]. To reduce any bias concerning perceptions of the robot's gender, we utilized the gender-neutral "Ivy" voice.

C. iSpy Testbed

The game incorporated into the study consists of the interaction partner prompting the NAO robot to identify colored blocks and is loosely based on the guessing game, *iSpy*. The NAO robot was positioned to face a small table holding the three, 1.5in blocks (green, red, and blue). The blocks were placed approximately 5in away from one another to prevent any misinterpretation of where the NAO was pointing. To initiate the game, the robot introduces itself and prompts the player to begin the round. Next, the human agent "spies" their desired block. Then, the robot successfully (or unsuccessfully) identifies the participant's selection. The human partner either

TABLE II
SAMPLE SCRIPT FOR THE CONTROL (NO-HUMOR) CONDITION.

NAO: *Let's begin round three.*
Human: *I spy something red.*
NAO: *Is it this one?* [gestures to the green block]
Human: *No.*
NAO: *Ok. Good game.*

affirms or refutes the robot's attempt. In the case of failure, the robot tries to recover with one of the four humor types or no humor. A sample script of the interaction is included in Table II and a visual of the interaction is included in Fig. 3.

V. STUDIES

A. Pilot Study

1) Design: For the main studies, we needed to first ensure that people found the jokes to be *funny*. To address this, we conducted a within-subjects pilot study ($n = 50$) via Amazon Mechanical Turk (MTurk) to obtain eight, funny jokes for the main study [72]. The experimenters drafted a set of 20 jokes (five jokes for each of the four humor categories) and collected funniness ratings for each joke.

2) Procedure: The participants were first instructed to review the electronic study information document for consent and the task instructions and to complete a brief demographic survey. Then, participants were shown 20 videos in a randomized order. In each clip, the robot misidentified a block and followed the mistake with a joke from one of the four humor categories. After viewing each interaction, the participants were asked to rate how funny the joke was on a Likert scale ranging from "not funny" (1) to "very funny" (5). For the recorded interactions, the longest video was 20 seconds, and the shortest was 12 seconds long ($M = 14$). At the conclusion of the study, participants were debriefed and compensated \$2 for participating in the study. Our pilot study was approved by the Institutional Review Board.

3) Participants: A total of 50 adults participated in the pilot study on MTurk (60% male ($n = 30$), 40% female ($n = 20$)). The mean age of participants was 33.76 years ($SD = 9.73$). All participants were based in the United States and required to be English speakers, at least 18 years of age or older, and have a Human Intelligence Task (HIT) approval rating of 99% or greater. We also added a timer to the video clips to ensure

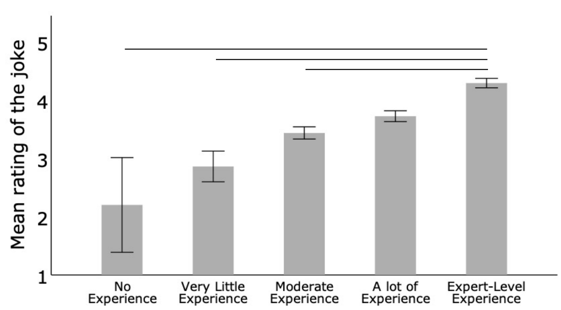

Fig. 2. Mean ratings on a scale from 1 (not funny) to 5 (very funny) based on previous experience with robots scale from 1 (none) to 5 (expert-level). Error bars 95% CI. Brackets are either significant at the .05, .01, or .001 level.

TABLE III
MEAN RATINGS FOR THE SELECTED JOKES ON A SCALE FROM 1 (NOT FUNNY) TO 5 (VERY FUNNY).

Affiliative	3.80
Aggressive	3.79
Self-defeating	3.68
Self-enhancing	3.78

that participants watched the videos in their entirety. The participants consisted of a wide variety of occupations, and the highest level of education mainly consisted of Bachelor's ($n = 36$) and Master's ($n = 11$) degrees. During the study participants also reported their experience with both NAO robots ($M = 3.38$, $SD = 1.23$) and robots in general ($M = 3.70$, $SD = 1.06$) on a Likert scale ranging from "no experience" (1) to "expert-level experience" (5).

4) Results: We took the mean rating of funniness for the 20 jokes. From there, we selected the two highest rated jokes for each humor category to use in our main study. As a result, we were left with a total of eight jokes to use in the main component of our study (Table III). Additionally, in the pilot study, we compared funniness ratings for the jokes across the four humor categories and examined how participant demographics affected these ratings. The results did not indicate a significant effect of age or gender on how the jokes were perceived. However, the results indicated a significant effect of previous robot experience on ratings of funniness $F(4, 45) = 9.22$, $p < .001$ (Fig. 2). In particular, the results indicated that participants with a higher experience with robots found the jokes to be funnier than participants with less experience. As a result, we chose to also consider previous robot experience as an independent variable in for the main study (RQ3 and RQ4).

B. Study 1: Dyadic Interaction

1) Design: For Study 1, we examined a dyadic, in-person human-robot interaction. We collected participant ratings of *warmth*, *competence*, and the robot as a *teammate*. For the *iSpy* setup, we added the eight best jokes from the pilot study to the Choregraphe program. In Choregraphe, we also utilized the speech recognition modules, which enabled the robot to identify to the three block colors upon instruction. In addition, we used the speech recognition modules to trigger humorous responses upon task failure.

Two NAO robots were used for the study, and participants were able to see the robots being switched between games. We also had the robots introduce themselves to further the distinction between the two. We chose the gender-neutral names Sam and Kris to avoid any potential bias in perceptions of gender. In addition, we modified the LED eye color of the two robots to blue and green, respectively.

The failure condition was programmed to always occur in the third round of Sam's interaction. However, the order that participants played against the two robots was randomized. Both robots utilized the same text-to-speech platform and condition as the robot in the pilot study. Both robots were programmed to operate autonomously; however, during the testing

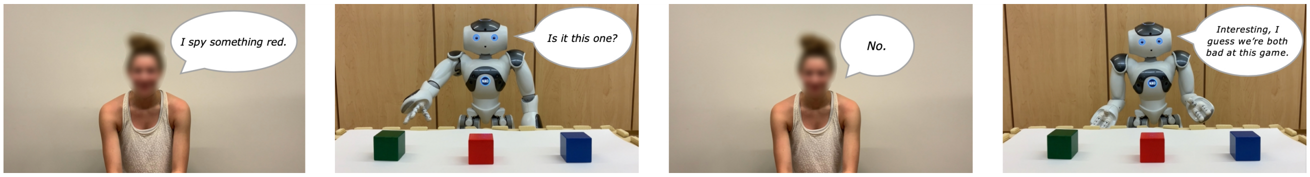

Fig. 3. A visual depiction of the interaction scenario between a dyadic participant and the NAO robot that uses aggressive humor upon failure. First, the participant indicates their desired block. Next, the robot attempts to identify the block. Then, the participant provides the robot with feedback on its selection. Finally, the robot attempts to recover from its failure with an aggressive joke.

stage, we found that the voice recognition module occasionally struggled to process heavy accents, poorly enunciated words, and quiet voices. As a result, we also included a Wizard-of-Oz method for progressing the interaction in cases where the robot's voice recognition module was unable to recognize the human participant's commands. During the study, one experimenter was on standby, outside of the participant's line-of-sight, to advance the NAO's actions when necessary.

2) Procedure: After reviewing the informed consent document and task instructions, participants were asked to complete a brief demographic survey. Next, participants were guided to a partitioned area (Fig. 1) and instructed to play the first game. Upon completion, an experimenter would instruct the participants to complete the first post-task survey. While participants completed the survey, another experimenter would swap out the robots. Then, participants were taken back to the partitioned area to play the second game with the second robot. Following this interaction, participants were asked to complete the final post-task survey. At the conclusion of the study, participants were debriefed and compensated. Additionally, we collected recordings of the dyadic participants playing the two games to use for the bystander condition on MTurk (See V-C).

3) Measures: To formulate our measures, we referred to the aforementioned literature on failure mitigation (See II-C). We selected measures that had been thoroughly explored in the context of HRI because we wanted to ground our observations on the effects of humor in cases of failure. Specifically, participants were asked to rate perceptions of the NAO robot's *warmth* and *competence*. For our third measure, *teammate*, participants were asked to provide subjective ratings for a relevant subset of perceptions which were then averaged to provide a holistic measure of perceptions of robot as a collaborative teammate. Specifically, participants were asked to reflect on whether the NAO robot was a good teammate, rate their willingness to work with the robot again, and express whether the robot met their expectations. All measures were rated across a 5-point Likert scale. We also collected measures of participant demographic information.

4) Participants: A total of 32 adults participated in the study (59.4% male ($n = 19$), 37.5% female ($n = 12$), 3.1% nonbinary ($n = 1$)). The mean age of participants was 24 years ($SD = 4.83$). All participants were required to be English speakers, available to participate in-person, and at least 18 years of age or older. The participants consisted of mostly graduate and undergraduate university students ($n = 16$ and $n = 15$, respectively). During the study, participants also reported their experience with both NAO robots ($M = 1.50$, $SD = 1.02$) and robots in general ($M = 2.38$, $SD = 1.04$). Participants were compensated with a $10 gift card for participating in the study. Our study was approved by the Institutional Review Board.

C. Study 2: Bystander Interaction

1) Design: In Study 2, we examined the bystander perspective in a human-robot interaction. Participants were instructed to watch and rate two games of the NAO robot playing the three rounds of *iSpy*. The study was conducted over MTurk to reach a large, diverse set of participants for the bystander condition. The recordings were presented to the MTurk participants in the same order that they were experienced by the in-person participants. In the third round of one of the games, the participants would watch the NAO robot misidentify one of the blocks and respond with either a joke from one of the four humor categories or no joke in the case of the control condition. We also collected demographic information at the start of each study to examine how these factors affected the different perceptions.

The recordings of the 32 interactions in Study 1 were shown to the MTurk participants for the bystander condition. Each interaction consisted of 2 videos: one with Sam and one with Kris. Personal identifiers were not included in the videos.

2) Procedure: After reviewing the electronic study information document for consent and the task instructions, participants were asked to complete a brief demographic survey. Next, the participants were shown one of the 32 previously recorded interactions. Each video was presented in the same order it was experienced by the in-person participants. After each video, participants were instructed to complete a post-task survey. At the conclusion of the study, participants were debriefed and compensated.

3) Measures: The Study 1 subjective measures on perceptions of warmth, competence, and the robot as a teammate as well as demographic measures were also used in Study 2.

4) Participants: A total of 256 adults participated in the study on MTurk (60.9% male ($n = 156$), 39.1% female ($n = 100$)). The mean age of participants was 35.7 years ($SD = 10.8$). All participants were based in the United States and required to be English speakers, at least 18 years of age or older, and have a HIT approval rating of 99% or greater. The participants consisted of a wide variety of occupations, and the highest level of education mainly consisted of Bachelor's ($n = 167$) and Master's ($n = 68$) degrees. During the study, participants also reported their experience with both NAO robots ($M = 3.52$, $SD = 1.33$) and robots in general

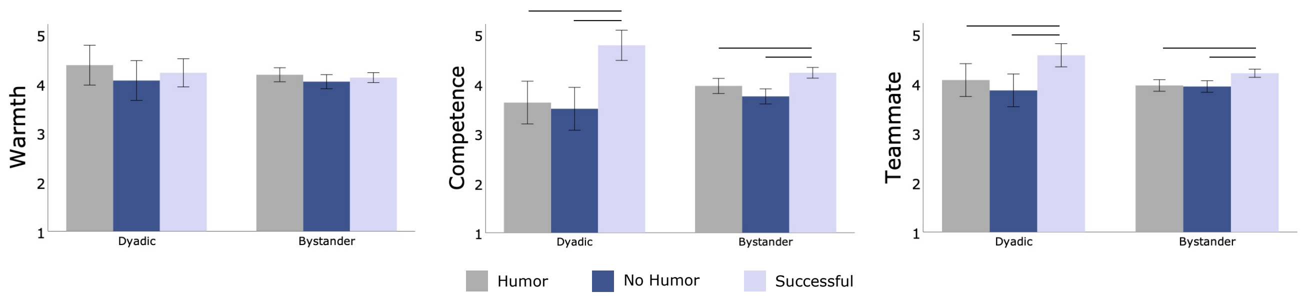

Fig. 4. Bar graphs depicting dyadic and bystander perceptions of warmth, competence, and the robot as a teammate for the robot that failed and used humor, failed and did not use humor, and successfully completed the task. Error bars 95% CI. Brackets are either significant at the .05, .01, or .001 level.

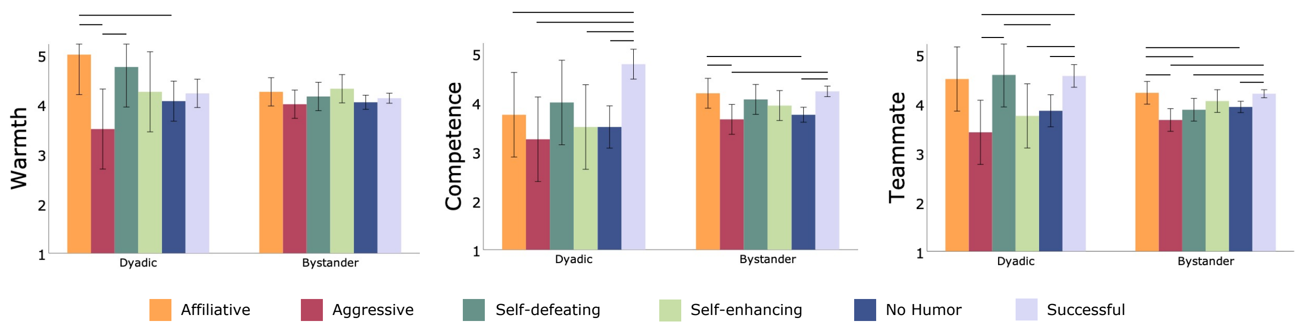

Fig. 5. Bar graphs showing dyadic and bystander perceptions of warmth, competence, and the robot as a teammate for the robot that failed and used one of the four humor types, failed and did not use humor, and successfully completed the task. Error bars 95% CI. Brackets are either significant at the .05, .01, or .001 level.

($M = 3.75$, $SD = 1.11$). Participants were compensated \$1 for participating in the study. Our study was approved by the Institutional Review Board.

VI. Results

To address our research questions, we evaluated the effects of humor, perspective, and prior robot experience on failure mitigation. Specifically, we examined perceptions of warmth, competence, and the robot as a teammate. For the analysis, inspection of the data and Levene's test provided no strong evidence against the assumption of constant variance. Additionally, we used Fisher's Least Significant Difference (LSD) test for post-hoc comparisons.

A. ***RQ1:*** *Effect of overall humor*

1) Warmth: We conducted a two-way ANOVA to examine the effect of general performance (successful task completion, failed task completion with humor, and failed task completion without humor) and perspective (dyadic vs. bystander) on perceptions of warmth. The results (Fig. 4) did not suggest an interaction between perspective and general performance on warmth ($F(2, 573) = .156$, $\eta^2 = .001$, $p = .856$). This indicates that the performance of the robot did not have significantly different effects on the ratings of warmth for the dyadic participants and bystanders. Furthermore, the dyadic participants did not differ significantly from bystanders in their ratings of warmth ($p = .360$).

2) Competence: To examine the effect of general performance and perspective on perceptions the robot's competence, we conducted a two-way ANOVA. The results (Fig. 4) suggest an interaction between perspective and general performance on competence ($F(2, 573) = 6.640$, $\eta^2 = .023$, $p = .001$). The results suggest that both the dyadic participants and the bystanders found the robot that did not experience failure to be significantly more competent than the humorous (dyadic ($p < .001$), bystander ($p = .005$)) and non-humorous (dyadic ($p < .001$), bystander ($p < .001$)) robots. Furthermore, the dyadic participants' ratings of competence were significantly higher than bystanders' ratings of competence for the successful robot ($p = .001$).

3) Teammate: Finally, to examine the effect of general performance and perspective on perceptions of the robot as a teammate, we conducted a two-way ANOVA. The results (Fig. 4) did not suggest an interaction between perspective and general performance on how the robot was perceived as a teammate ($F(2, 573) = 2.162$, $\eta^2 = .008$, $p = .116$). However, the results indicate that the dyadic participants found the successful robot to be a better teammate than the bystander participants rated ($p < .001$).

B. ***RQ2:*** *Effect of humor type*

1) Warmth: We conducted a two-way ANOVA, to examine the effect of type-specific performance, and perspective on perceptions of warmth. The results (Fig. 5) did not suggest an interaction between perspective and type-specific performance on warmth ($F(5, 570) = 1.135$, $\eta^2 = .01$, $p = .340$). Meaning, the dyadic participants did not differ significantly from bystanders in their ratings of warmth for the type-specific performance scenarios. Furthermore, we did not find evidence that the dyadic participants differed significantly from bystanders in their ratings of warmth ($p = .322$).

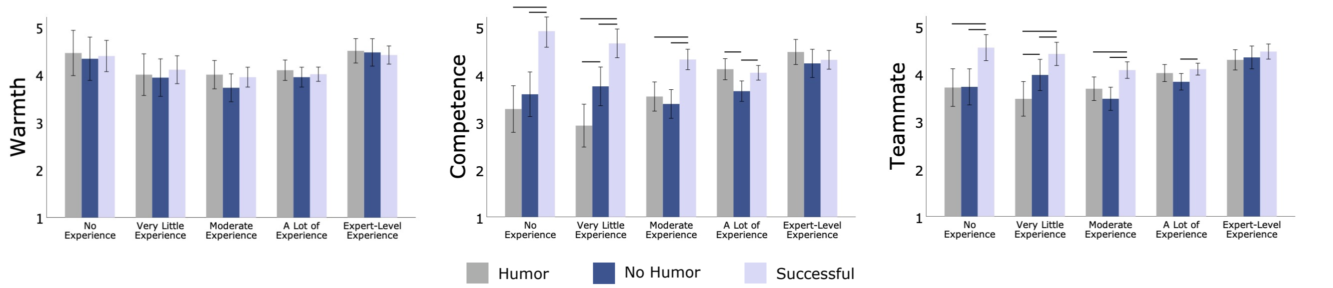

Fig. 6. Bar graphs depicting perceptions of warmth, competence, and the robot as a teammate for the different participant robot experience levels. The plots show results for the robot that failed and used humor, failed and did not use humor, and successfully completed the task. Error bars 95% CI. Brackets are either significant at the .05, .01, or .001 level.

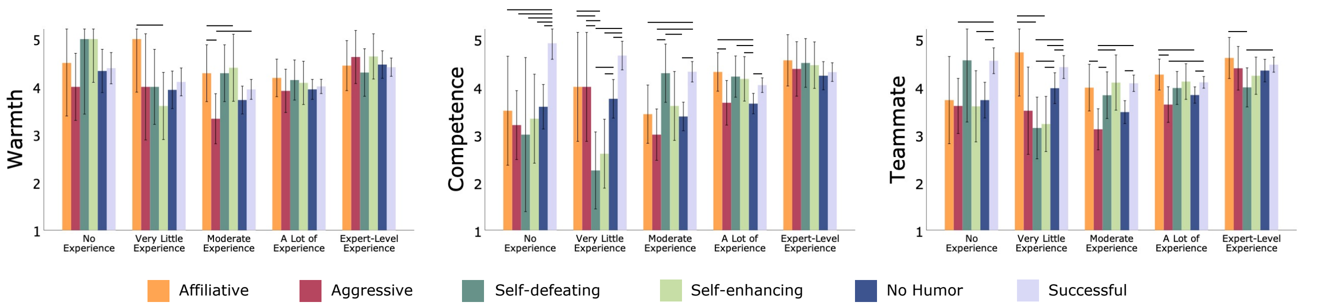

Fig. 7. Bar graphs showing perceptions of warmth, competence, and the robot as a teammate for the previous robot experience levels. The plots illustrate results for the robot that failed and used one of the four humor types, failed and did not use humor, and successfully completed the task. Error bars 95% CI. Brackets are either significant at the .05, .01, or .001 level.

2) Competence: To examine the effect of type-specific performance and perspective on perceptions of competence, we conducted a two-way ANOVA. The results (Fig. 5) suggest an interaction between perspective and type-specific performance on competence ($F(5, 570) = 2.758$, $\eta^2 = .024$, $p = .018$). In particular, the dyadic participants differed significantly from bystanders in their ratings of competence for the type-specific performance scenarios. Specifically, the dyadic participants found the robot that successfully completed the task to be significantly more competent than the robots that used affiliative ($p = .027$), aggressive ($p = .001$), self-enhancing ($p = .006$), or no-humor ($p < .001$). Contrarily, the bystanders found the affiliative robot and the successful robot to be significantly more competent than the robots that used aggressive (affiliative ($p = .016$), successful ($p = .001$)) or no-humor (affiliative ($p = .012$), successful ($p < .001$)). Additionally, the results suggest that the dyadic participants perceived the successful robot as significantly more competent than the bystander participants ($p = .001$).

3) Teammate: We conducted a two-way ANOVA to examine the effect of type-specific performance and perspective on perceptions of the robot as a teammate. The results (Fig. 5) did not suggest an interaction between perspective and type-specific performance on perceptions of the robot as a teammate ($F(5, 570) = 1.996$, $\eta^2 = .017$, $p = .078$). However, the dyadic participants had significantly higher ratings of the robot as a teammate than the bystanders for the robot that used self-defeating humor ($p = .011$). Furthermore, the dyadic participants had significantly higher ratings of the robot as a teammate than the bystanders for the successful robot ($p < .001$).

C. ***RQ3:*** *Effect of overall humor and robot experience*

We conducted a series of two-way ANOVAs to examine the effect of general performance and previous experience with a robot on perceptions of warmth, competence, and potential as a teammate. We did not consider perspective as a factor in this analysis because the number of participants in the dyadic study did not indicate the full spectrum of robot experience to perform comprehensive significance testing.

1) Warmth: The results (Fig. 6) did not suggest an interaction between previous robot experience and general performance on warmth ($F(8, 567) = .213$, $\eta^2 = .003$, $p = .989$). This indicates that the performance of the robot did not have significantly different effects on the ratings of warmth for participants of varying experience levels.

2) Competence: The results (Fig. 6) suggest an interaction between previous robot experience and general performance on competence ($F(8, 567) = 9.196$, $\eta^2 = .116$, $p < .001$). In particular, participants with moderate to no previous robot experience found the successful robot to be significantly more competent ($p < .001$) than the robots that failed (both humorous and non-humorous).

3) Teammate: The results (Fig. 6) also suggested an interaction between perspective and general performance on how the robot was perceived as a teammate ($F(8, 567) = 3.061$, $\eta^2 = .002$, $p = .042$). For instance, participants with less experience (no to moderate) found the successful robot to be significantly better than the failing robot (for participants with no experience with humorous robot ($p = .001$) and with non-humorous robot ($p = .001$), and for participants with moderate experience with humorous robot ($p = .011$) and with non-humorous robot ($p < .001$)). Contrarily, participants with expert-levels of previous robot experience did not have significant differences in ratings of the robot as a teammate (humorous robot ($p = .197$), non-humorous robot ($p = .400$)). The results also suggest that participants with expert level experience found the humorous robot to be a significantly better teammate than the participants with less experience (no experience ($p = .011$), moderate experience ($p < .001$)).

D. ***RQ4:*** *Effect of humor type and robot experience*

To examine the effect of type-specific performance and previous experience with robot on perceptions of the robot's warmth, competence, and potential as a teammate, we conducted a series of two-way ANOVAs. Similar to RQ3, we did not consider perspective as a factor in this analysis as the number of participants in the dyadic study did not indicate the full spectrum of robot experience to perform comprehensive significance testing.

1) Warmth: The results (Fig. 7) did not suggest an interaction between previous robot experience and type-specific performance on warmth ($F(5, 570) = .781$, $\eta^2 = .028$, $p = .738$). This indicates that the type-specific performance of the robot did not have significantly different effects on the ratings of warmth for participants of varying experience levels.

2) Competence: The results (Fig. 7) suggest an interaction between previous robot experience and type-specific performance on competence ($F(5, 570) = 4.574$, $\eta^2 = .143$, $p < .001$). Specifically, participants with less previous robot experience found the successful robot to be significantly more competent than the robots that failed and used no humor ($p < .001$) or one of the four humor types (affiliative ($p = .02$), aggressive ($p < .001$), self-enhancing ($p = .002$), self-defeating ($p = .023$)).

3) Teammate: Finally, the results (Fig. 7) also suggest an interaction between perspective and type-specific performance on how the robot was perceived as a teammate ($F(5, 570) = 3.061$, $\eta^2 = 11.009$, $p = .005$). Participants with moderate to no previous robot experience were more diversified in their ratings of the robot as a teammate for the type-specific performance. Whereas, participants with high levels of previous robot experience had less variation in their ratings of the robot as a teammate.

VII. Discussion

In our study, we examined the effects of humor as a failure recovery method on perceptions of the robot partner. Specifically, we evaluated the effects of performance (successful vs. fail), overall humor (humor vs. no-humor), humor type, perspective (dyadic vs. bystander), and previous robot experience (no to expert-level experience) on ratings of warmth, competence, and the robot as a teammate. For **RQ1**, we did not observe a significant interaction between humor and perspective on perceptions of the robot's competence. However, the results suggest that the dyadic participants found the successful robot to be significantly more competent than the bystander participants. Additionally, we observed a significant interaction between humor type and perspective on ratings of competence (**RQ2**). In particular, the bystanders found the successful robot and the robot that failed and used affiliative humor to be more competent that the failing robots that used no or aggressive humor; whereas, the dyadic participants found the successful robot to be significantly more competent than the robots that failed and used affiliative, aggressive, self-enhancing, or no-humor. Moreover, the dyadic participants significantly differed from bystanders in their ratings of the successful robot's competence.

The results also indicate that dyadic participants found the robot that used self-defeating humor to be a better teammate than the bystanders observed. We reason that, for a direct interaction partner, self-defeating humor is a better recovery method for the robot's potential as a teammate than for bystanders. It could be that the vulnerability expressed in self-defeating humor is more endearing in intimate, dyadic interactions than in more observational, bystander interactions.

For **RQ3** and **RQ4**, we explored the effect of previous robot experience, humor, and performance on ratings of warmth, competence, and teammate. The results suggest a significant interaction between performance and ratings of the robot as a teammate. Overall, participants with moderate to no previous robot experience found the successful robot to be a better teammate than the failing humorous and non-humorous robots. However, when isolating humor type, the results suggest that participants with no experience and moderate experience are more diversified in their perceptions of the robot as a teammate. Contrarily, the results indicate that participants with high levels of previous robot experience did not significantly differ in ratings of the robot as a teammate. We posit that participants with high levels of previous robot experience are more familiar with the robot's performance potential and capabilities and did not find the robot's task failure to be as much of a negative reflection of its potential as a teammate.

VIII. Conclusion and Future Work

In this paper, we have presented a human-robot interaction study to examine the effect of advanced, more human-like humor strategies to mitigate task failure employed by a robot. We aimed to observe if the humorous failing robot would be considered as warm, as competent, and as good of a teammate as the successful robot by examining the three conditions (successful, fail with humor, and fail without humor). The ratings of warmth, competence, and teammate for the successful robot illustrate that performance still greatly affects perceptions of the robot partner. While certain types of humor can minimize the effects of failure on perceptions of warmth, competency, and the robot as a teammate, it cannot completely negate them.

In the future, we plan to examine how humor as a failure recovery strategy compares to other methods (explanations, apologies, and preemptive attempts to set expectations). We also plan to examine the interplay between humor, failure severity, and trust. Specifically, we are interested in observing how more advanced humor affects trust levels when used in response to failures of varying severity.

REFERENCES

[1] M. C. Gombolay, X. J. Yang, B. Hayes, N. Seo, Z. Liu, S. Wadhwania, T. Yu, N. Shah, T. Golen, and J. A. Shah, "Robotic assistance in coordination of patient care." in *Robotics: Science and Systems*, 2016.

[2] J. Wirtz, P. G. Patterson, W. H. Kunz, T. Gruber, V. N. Lu, S. Paluch, and A. Martins, "Brave new world: Service robots in the frontline," *Journal of Service Management*, vol. 29, no. 5, pp. 907–931, 11 2018.

[3] I. Leite, C. Martinho, and A. Paiva, "Social robots for long-term interaction: A survey," *International Journal of Social Robotics*, vol. 5, no. 2, pp. 291–308, 2013.

[4] T. Belpaeme, J. Kennedy, A. Ramachandran, B. Scassellati, and F. Tanaka, "Social robots for education: A review," *Science Robotics*, vol. 3, no. 21, 2018.

[5] A. Kubota, E. I. Peterson, V. Rajendren, H. Kress-Gazit, and L. D. Riek, "Jessie: Synthesizing social robot behaviors for personalized neurorehabilitation and beyond," in *Proceedings of the 2020 ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, ser. HRI '20, 2020, pp. 121–130.

[6] S. Valencia, M. Luria, A. Pavel, J. P. Bigham, and H. Admoni, "Co-designing socially assistive sidekicks for motion-based aac," in *Proceedings of the 2021 ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, ser. HRI '21. Association for Computing Machinery, 2021, p. 24–33.

[7] T. Iqbal and L. D. Riek, "Human robot teaming: Approaches from joint action and dynamical systems," *Humanoid Robotics: A Reference*, pp. 2293–2312, 2017.

[8] S. W. Logan, H. A. Feldner, K. R. Bogart, B. Goodwin, S. M. Ross, M. A. Catena, A. A. Whitesell, Z. J. Sefton, W. D. Smart, and J. C. Galloway, "Toy-based technologies for children with disabilities simultaneously supporting self-directed mobility, participation, and function: A tech report," *Frontiers in Robotics and AI*, vol. 4, p. 7, 2017.

[9] N. Randall, C. C. Bennett, S. Šabanović, S. Nagata, L. Eldridge, S. Collins, and J. A. Piatt, "More than just friends: in-home use and design recommendations for sensing socially assistive robots (sars) by older adults with depression," *Paladyn, Journal of Behavioral Robotics*, vol. 10, no. 1, pp. 237–255, 2019.

[10] S. Chernova and A. L. Thomaz, "Robot learning from human teachers," *Synthesis Lectures on Artificial Intelligence and Machine Learning*, vol. 8, no. 3, pp. 1–121, 2014.

[11] M. Natarajan and M. Gombolay, *Effects of Anthropomorphism and Accountability on Trust in Human Robot Interaction*, ser. HRI '20. Association for Computing Machinery, 2020, p. 33–42.

[12] T. Iqbal, S. Rack, and L. D. Riek, "Movement coordination in human–robot teams: A dynamical systems approach," *IEEE Transactions on Robotics*, vol. 32, no. 4, pp. 909–919, 2016.

[13] T. Iqbal, S. Li, C. Fourie, B. Hayes, and J. A. Shah, "Fast Online Segmentation of Activities from Partial Trajectories," in *IEEE International Conference on Robotics and Automation (ICRA)*, 2019.

[14] S. Javdani, H. Admoni, S. Pellegrinelli, S. S. Srinivasa, and J. A. Bagnell, "Shared autonomy via hindsight optimization for teleoperation and teaming," *The International Journal of Robotics Research*, vol. 37, no. 7, pp. 717–742, 2018.

[15] H. Admoni, T. Weng, B. Hayes, and B. Scassellati, "Robot nonverbal behavior improves task performance in difficult collaborations," in *2016 11th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, 2016, pp. 51–58.

[16] A. Saran, S. Majumdar, E. S. Short, A. Thomaz, and S. Niekum, "Human gaze following for human-robot interaction," in *2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2018, pp. 8615–8621.

[17] M. M. Islam and T. Iqbal, "Hamlet: A hierarchical multimodal attention-based human activity recognition algorithm," in *2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2020, pp. 10 285–10 292.

[18] ——, "Multi-gat: A graphical attention-based hierarchical multimodal representation learning approach for human activity recognition," *IEEE Robotics and Automation Letters (RA-L)*, 2021.

[19] M. L. Traeger, S. Strohkorb Sebo, M. Jung, B. Scassellati, and N. A. Christakis, "Vulnerable robots positively shape human conversational dynamics in a human–robot team," *Proceedings of the National Academy of Sciences*, vol. 117, no. 12, pp. 6370–6375, 2020.

[20] D. J. Brooks, D. J. Curtin, J. T. Kuczynski, J. J. Rodriguez, A. Steinfeld, and H. A. Yanco, "A communication paradigm for human-robot interaction during robot failure scenarios," in *Human-Machine Shared Contexts*, W. F. Lawless, R. Mittu, and D. A. Sofge, Eds. Academic Press, 2020, pp. 277–306.

[21] S. Tellex, R. Knepper, A. Li, D. Rus, and N. Roy, "Asking for help using inverse semantics," in *Proceedings of Robotics: Science and Systems*, 2014.

[22] T. Williams, P. Briggs, and M. Scheutz, "Covert robot-robot communication: Human perceptions and implications for human-robot interaction," *Journal of Human-Robot Interaction*, vol. 4, no. 2, p. 24–49, 2015.

[23] A. S. Clair and M. Matarić, "How robot verbal feedback can improve team performance in human-robot task collaborations," in *2015 10th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, ser. HRI '15, 2015, pp. 213–220.

[24] N. Martelaro, V. C. Nneji, W. Ju, and P. Hinds, "Tell me more designing hri to encourage more trust, disclosure, and companionship," in *2016 11th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, ser. HRI '16, 2016, pp. 181–188.

[25] L. Takayama, V. Groom, and C. Nass, *I'm Sorry, Dave: I'm Afraid i Won't Do That: Social Aspects of Human-Agent Conflict*. Association for Computing Machinery, 2009, p. 2099–2108.

[26] F. Correia, S. Mascarenhas, R. Prada, F. S. Melo, and A. Paiva, "Group-based emotions in teams of humans and robots," in *Proceedings of the 2018 ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, ser. HRI '18. Association for Computing Machinery, 2018, p. 261–269.

[27] Z. Han, E. Phillips, and H. A. Yanco, "The need for verbal robot explanations and how people would like a robot to explain itself," *Journal of Human-Robot Interaction*, vol. 10, no. 4, 09 2021.

[28] S. Kobel and A. Groeppel-Klein, "No laughing matter, or a secret weapon? exploring the effect of humor in service failure situations," *Journal of Business Research*, vol. 132, pp. 260–269, 2021.

[29] N. R. Norrick and A. Spitz, "Humor as a resource for mitigating conflict in interaction," *Journal of Pragmatics*, vol. 40, no. 10, pp. 1661–1686, 2008.

[30] N. Mirnig, G. Stollnberger, M. Giuliani, and M. Tscheligi, "Elements of humor: How humans perceive verbal and non-verbal aspects of humorous robot behavior," in *Proceedings of the Companion of the 2017 ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, ser. HRI '17. Association for Computing Machinery, 2017, p. 211–212.

[31] I. M. Menne, B. P. Lange, and D. C. Unz, "My humorous robot: Effects of a robot telling jokes on perceived intelligence and liking," in *Companion of the 2018 ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, ser. HRI '18. Association for Computing Machinery, 2018, p. 193–194.

[32] M. Vázquez, A. May, A. Steinfeld, and W.-H. Chen, "A deceptive robot referee in a multiplayer gaming environment," in *2011 International Conference on Collaboration Technologies and Systems (CTS)*, 2011, pp. 204–211.

[33] I. Leite, M. McCoy, M. Lohani, N. Salomons, K. McElvaine, C. Stokes, S. Rivers, and B. Scassellati, "Autonomous disengagement classification and repair in multiparty child-robot interaction," in *2016 25th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*, 2016, pp. 525–532.

[34] M. F. Jung, N. Martelaro, and P. J. Hinds, "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*, ser. HRI '15. Association for Computing Machinery, 2015, p. 229–236.

[35] W. J. Smith, K. V. Harrington, and C. P. Neck, "Resolving conflict with humor in a diversity context," *Journal of Managerial Psychology*, vol. 15, no. 6, p. 606–625, 2000.

[36] S. F. Dziegielewski, "Humor," *International Journal of Mental Health*, vol. 32, no. 3, pp. 74–90, 2003.

[37] M. H. Abel, "Humor, stress, and coping strategies," *Humor: International Journal of Humor Research*, vol. 15, no. 4, pp. 365–381, 2002.

[38] J.-Y. Mao, J. T.-J. Chiang, Y. Zhang, and M. Gao, "Humor as a relationship lubricant: The implications of leader humor on transformational leadership perceptions and team performance," *Journal of Leadership & Organizational Studies*, vol. 24, no. 4, pp. 494–506, 2017.

[39] B. M. Savage, H. L. Lujan, R. R. Thipparthi, and S. E. DiCarlo, "Humor, laughter, learning, and health! a brief review," *Advances in Physiology Education*, vol. 41, no. 3, pp. 341–347, 2017.

[40] J. Mesmer-Magnus, D. J. Glew, and C. Viswesvaran, "A meta-analysis of positive humor in the workplace," *Journal of Managerial Psychology*, vol. 27, no. 2, pp. 155–190, 02 2012.

[41] A. M. Bippus, "Humor motives, qualities, and reactions in recalled conflict episodes," *Western Journal of Communication*, vol. 67, no. 4, pp. 413–426, 2003.

[42] C. Warren, A. Barsky, and A. P. McGraw, "Humor, comedy, and consumer behavior," *Journal of Consumer Research*, vol. 45, no. 3, pp. 529–552, 03 2018.

[43] J. Vilk and N. T. Fitter, "Comedians in cafes getting data: Evaluating timing and adaptivity in real-world robot comedy performance," in *Proceedings of the 2020 ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, ser. HRI '20. Association for Computing Machinery, 2020, p. 223–231.

[44] ——, "Comedy by jon the robot," in *Companion of the 2020 ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, ser. HRI '20. Association for Computing Machinery, 2020, pp. 223–231.

[45] ——, "Jon the robot goes hollywood," in *Companion of the 2020 ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, ser. HRI '20. Association for Computing Machinery, 2020, p. 644.

[46] A. Nijholt, "Robotic stand-up comedy: State-of-the-art," in *Distributed, Ambient and Pervasive Interactions: Understanding Humans (DAPI)*. Springer International Publishing, 07 2018, pp. 391–410.

[47] K. Weber, H. Ritschel, I. Aslan, F. Lingenfelser, and E. André, "How to shape the humor of a robot - social behavior adaptation based on reinforcement learning," in *Proceedings of the 20th ACM International Conference on Multimodal Interaction (ICMI)*, ser. ICMI '18. Association for Computing Machinery, 2018, p. 154–162.

[48] B. Stoll, M. F. Jung, and S. R. Fussell, "Keeping it light: Perceptions of humor styles in robot-mediated conflict," in *Companion of the 2018 ACM/IEEE International Conference on Human-Robot Interaction*, ser. HRI '18. Association for Computing Machinery, 2018, p. 247–248.

[49] M. F. Jung, "Affective grounding in human-robot interaction," in *2017 12th ACM/IEEE International Conference on Human-Robot Interaction (HRI*, 2017, pp. 263–273.

[50] E. J. de Visser, M. M. Peeters, M. F. Jung, S. Kohn, T. H. Shaw, R. Pak, and M. A. Neerincx, "Towards a theory of longitudinal trust calibration in human–robot teams," *International Journal of Social Robotics*, vol. 12, p. 459–478, 2020.

[51] R. Oliveira, P. Arriaga, M. Axelsson, and A. Paiva, "Humor-robot interaction: A scoping review of the literature and future directions," *International Journal of Social Robotics*, 2020.

[52] S. Strohkorb Sebo, M. Traeger, M. Jung, and B. Scassellati, "The ripple effects of vulnerability: The effects of a robot's vulnerable behavior on trust in human-robot teams," in *Proceedings of the 2018 ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, ser. HRI '18. Association for Computing Machinery, 2018, p. 178–186.

[53] T. Adamson, C. B. Lyng-Olsen, K. Umstattd, and M. Vázquez, "Designing social interactions with a humorous robot photographer," in *Proceedings of the 2020 ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, ser. HRI '20. Association for Computing Machinery, 2020, p. 233–241.

[54] M. Giuliani, N. Mirnig, G. Stollnberger, S. Stadler, R. Buchner, and M. Tscheligi, "Systematic analysis of video data from different human–robot interaction studies: a categorization of social signals during error situations," *Frontiers in Psychology*, vol. 6, no. 931, 2015.

[55] S. Honig and T. Oron-Gilad, "Understanding and resolving failures in human-robot interaction: Literature review and model development," *Frontiers in Psychology*, vol. 9, p. 861, 2018.

[56] S. Reig, E. J. Carter, T. Fong, J. Forlizzi, and A. Steinfeld, "Flailing, hailing, prevailing: Perceptions of multi-robot failure recovery strategies," in *Proceedings of the 2021 ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, ser. HRI '21. Association for Computing Machinery, 2021, p. 158–167.

[57] D. Cameron, S. de Saille, E. C. Collins, J. M. Aitken, H. Cheung, A. Chua, E. J. Loh, and J. Law, "The effect of social-cognitive recovery strategies on likability, capability and trust in social robots," *Computers in Human Behavior*, vol. 114, p. 106561, 2021.

[58] G. M. Lucas, J. Boberg, D. Traum, R. Artstein, J. Gratch, A. Gainer, E. Johnson, A. Leuski, and M. Nakano, "Getting to know each other: The role of social dialogue in recovery from errors in social robots," in *Proceedings of the 2018 ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, ser. HRI '18. Association for Computing Machinery, 2018, p. 344–351.

[59] S. Choi, A. S. Mattila, and L. E. Bolton, "To err is human(-oid): How do consumers react to robot service failure and recovery?" *Journal of Service Research*, vol. 24, no. 3, pp. 354–371, 2021.

[60] P. Robinette, W. Li, R. Allen, A. M. Howard, and A. R. Wagner, "Overtrust of robots in emergency evacuation scenarios," in *2016 11th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, 2016, pp. 101–108.

[61] M. Salem, G. Lakatos, F. Amirabdollahian, and K. Dautenhahn, "Would you trust a (faulty) robot? effects of error, task type and personality on human-robot cooperation and trust," in *Proceedings of the Tenth Annual ACM/IEEE International Conference on Human-Robot Interaction*. Association for Computing Machinery, 2015, p. 141–148.

[62] M. R. Fraune, Y. Nishiwaki, S. Sabanović, E. R. Smith, and M. Okada, "Threatening flocks and mindful snowflakes: How group entitativity affects perceptions of robots," in *Proceedings of the 2017 ACM/IEEE International Conference on Human-Robot Interaction*, ser. HRI '17. Association for Computing Machinery, 2017, p. 205–213.

[63] T. Iqbal and L. D. Riek, "A method for automatic detection of psychomotor entrainment," *IEEE Transactions on Affective Computing*, vol. 7, no. 1, pp. 3–16, 2016.

[64] M. S. Yasar and T. Iqbal, "A scalable approach to predict multi-agent motion for human-robot collaboration," *IEEE Robotics and Automation Letters*, vol. 6, no. 2, pp. 1686–1693, 2021.

[65] ——, "Improving human motion prediction through continual learning," *ACM/IEEE Int. Conf. on Human-Robot Interaction (HRI), LEAP-HRI Workshop*, 2021.

[66] T. Iqbal and L. D. Riek, "Coordination dynamics in multihuman multirobot teams," *IEEE Robotics and Automation Letters*, vol. 2, no. 3, pp. 1712–1717, 2017.

[67] M. R. Fraune, S. Šabanović, and E. R. Smith, "Some are more equal than others: Ingroup robots gain some but not all benefits of team membership," *Interaction Studies*, vol. 21, no. 3, pp. 303–328, 2020.

[68] E. Short and M. J. Mataric, "Robot moderation of a collaborative game: Towards socially assistive robotics in group interactions," in *2017 26th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*, 2017, pp. 385–390.

[69] R. A. Martin, P. Puhlik-Doris, G. Larsen, J. Gray, and K. Weir, "Individual differences in uses of humor and their relation to psychological well-being: Development of the humor styles questionnaire," *Journal of Research in Personality*, vol. 37, no. 1, pp. 48–75, 2003.

[70] SoftBank Robotics, "Nao," https://www.softbankrobotics.com/emea/en/nao, 2021, retrieved: July 29, 2021.

[71] Amazon Web Services, Inc, "Amazon polly," https://aws.amazon.com/polly/, 2021, retrieved: June, 2021.

[72] H. N. Green, M. M. Islam, S. Ali, and T. Iqbal, "ispy a humorous robot: Evaluating the perceptions of humor types in a robot partner," in *Proceedings of AAAI Spring Symposium on Putting AI in the Critical Loop: Assured Trust and Autonomy in Human-Machine Teams*, 2022.