No Name, No Voice, Less Trust: Robot Group Identity Performance, Entitativity, and Trust Distribution

Alexandra Bejarano and Tom Williams

***Abstract*— Human interactions with robot groups are more complex than interactions with individual robots. This is especially true for groups of robots that do not have humanlike 1-1 associations between bodies and identities, such as when multiple robots share a single identity. This is further complicated by the lack of direct observability of the relationship between body and identity, which may be inferred by users on the basis of various robot group identity performance strategies. Previous research on *Deconstructed Trustee Theory* has argued that this complexity is critical, as different perceived body-identity configurations may lead users to build and develop trust in distinct ways. In this paper, we thus investigate (n=94) the ways that different robot group identity performance strategies might influence the distribution of trust amongst robot group members, as well as the impact of these strategies on perceptions of robot group entitativity.**

Fig. 1: Screenshot of study video shown to participants. A human interacts with 3 robots at an airport help desk.

Introduction

Human-robot interaction (HRI) research often investigates dyadic interactions; but non-dyadic interactions are becoming increasingly common, such as in healthcare [1], industrial [2], and educational [3] contexts. As such, it is critical to understand how human observers and interactants mentally model robot groups, and how those mental models may depend on different features of robots' designs. This poses a unique challenge for the HRI community: While over a century of research has been performed in fields like social psychology and sociology to characterize how people organize into, behave in, and perceive human groups [4], [5], [6], robot groups may manifest in new, and distinctly non-humanlike ways not captured by that literature.

Specifically, *robot group identity performance* [7] may deviate from humanlike presentations in the performed association between bodies (physical constructs), identities (performed personas), and minds (cognitive systems). In humans, this association is necessarily 1-1-1, where one mind is associated with one body and performs one identity. In contrast, identity within a robot group may depend on *group identity observables* (design cues like names and speech behavior [8]) that mediate perceived associations between minds, bodies, and identities. For instance, a robot group that shares a name and voice may be perceived as sharing an identity and mind across multiple bodies. This may influence in-group dynamics constructs like entitativity [9] and interpersonal psychological constructs like trust [10], thus influencing the overall dynamics and quality of interactions.

Previous research suggests that robot group identity performance may affect these factors. Bejarano et al. [7] hypothesized that when multiple bodies share an identity, they are perceived as being controlled by a single mind, and as highly entitative. Meanwhile, Williams et al. [11] argue that the number of bodies and identities within users' mental model may dictate where and how they believe trust can be placed.

In this work, we build on the work of Bejarano et al. [7] and Williams et al. [11]' notion of *trust loci*, to interrogate the relationship between robot group identity performance, entitativity, and trust. Through a human-subject study (n=94), we explore (using an experimental context with physical robots) how group identity performance strategy influences perceived group entitativity and human-robot trust.

Related Work

In this section, we briefly review the literature on human perception of and interaction with robot groups, in order to motivate our specific research questions and hypotheses.

Interactions with Groups

Previous work has shown impacts of robot group cardinality on interaction [12], [13], [14], [15], [16]. While HRI research often investigates dyadic interactions, there has been increasing work focusing on non-dyadic interactions [17]. However, when group interactions are investigated in HRI, it is typically in the context of one robot interacting with a group of humans, or in the context of one or more humans interacting with a "group" of only two robots [18]. In contrast, previous research has shown that it may take three or more similar robots to be perceived as a social group [19]. As such, in this paper, we focus on interactions with groups of three robots.

Work on robot groups of this size includes research on interactions with very large robot groups (e.g. swarms [20],

The authors are with the MIRRORLab at the Colorado School of Mines, Golden, CO, USA {abejarano,twilliams}@mines.edu

[21]), and on the control of multi-robot systems (e.g. teleoperating furniture robots [22] and social robots [23], [24]). Yet there is little work on *direct verbal* interactions with robot groups, where the robots are not simply meant to be controlled by a human operator. Thus, we examine interactions with robots that are actively involved and use natural language to communicate with human interactants.

Furthermore, robot group cardinality is especially important due to the way it mediates other dimensions of human-robot interaction. Fraune et al. [12] investigated human interactions with robot groups, and found interactions between the number of robots in a group and robot morphology on key HRI measures. Specifically, Fraune et al. found that anthropomorphic robot groups led to more positive perceptions (e.g. likability) and mechanomorphic robot groups led to more negative perceptions, compared to individual robots of the same type. Further, Fraune et al. [25] found that people interacted more with robots in groups than individual robots, and that robots were rated more positively in a group than alone when robots displayed task-oriented behavior, but not when they displayed social behavior. Hence, perceptions of and interactions with robots are not only influenced by the number of robots present, but also by how those robots are presented, especially in terms of their appearance and behavior [18], [26], [27], [14], [28], [29], [7], [30], [11], [31], [32]. These findings, however, are mainly focused on perceptions of individual robots within groups, or perceptions of interactions themselves. In contrast, we are interested in how different design choices might lead to differences in how people mentally model groups of robots themselves.

Mental Modeling of Groups of Robots

We argue that the choices robot designers make about robot groups, especially regarding how those robots present themselves, will shape the mental models users develop for groups of robots, with implications for group dynamics like entitativity and interpersonal constructs like trust.

Group Dynamics and Identity Performance Strategies: Abrams et al. [9] outline a framework of three group dynamics concepts (ingroup identitification, cohesion, and entitativity) which play a key role in robot group perception. We will primarily focus on perceived entitativity because of our interest in overall group perception (cf. [19], [27], [26]). Entitativity can be manipulated through changes in a group's behavior or appearance. Nawroj et al. [30] explored factors that may influence perceived entitativity, finding that behavioral mimicry has a significant influence while appearance similarity and gaze direction had minor influence on perceived entitativity. The perceived entitativity of a group can then influence further human perceptions of the group. For instance, Fraune et al. [27] found that in cooperative HRI, perceptions of entitativity increased positive perceptions of robots and willingness to interact with those robots.

Yet notions of entitativity become complicated by non-humanlike dimensions of robot groups. Robot groups need not have the 1-1-1 association of mind-body-identity that humans do, and may change how bodies and identities are associated on-the-fly. In other words, robot groups can be designed to use *identity performance strategies* in which the presentation and distribution of robot identity is manipulated through changes in identity observables (e.g. how names and voices are assigned among a group). Different identity performance strategies may then evoke different user mental models of the group and its members [8], [33], [7]. Bejarano et al. [7], Luria et al. [34], and Reig et al. [35] discuss a variety of robot group identity performance strategies, such as (1) One-for-one, where each robot body is inhabited by one mind/identity; (2) One-for-all, where one mind/identity inhabits multiple bodies simultaneously; (3) Re-embodiment, where one mind/identity can move from one body to another; and (4) Co-embodiment, where multiple minds/identities inhabit a single body. Critically, these are *performance strategies* because the mind-body-identity alignment inferred by observers from various cues (i.e., the observables at their level of abstraction [8]) need not reflect the robots'"true" cognitive and software architectural structure. As such, robot group identity performance strategy complicates reasoning about the number of "robots" in an interaction and the way entitativity between robots can and should be assessed.

Consequently, in a broad sense-making study using illustrated storyboards, Bejarano et al. [7] examined the relationship between robot group identity performance strategies and mental model formation. They demonstrated how group identity observables lead observers to develop different mental models of robot groups in regards to the group's intelligence distribution and social relationships, as well as the group's entitativity. Specifically, Bejarano et al. [7] found that a group may present itself as all robot bodies having a shared identity to evoke perceptions of high entitativity and a user mental model where all robot bodies are perceived to be controlled by a single mind [7]. Similarly, Reig et al. [36] with short vignettes found that devices perceived as being similar or devices that act on commands given to others are more likely perceived as being one system. In this work, we further examine the relationship between identity performance strategies and entitativity, but instead use videos of real robots to gauge human perceptions.

Interpersonal Psychological Constructs and Identity Performance Strategies: The distinction between robot group identity performance strategies is important, as different strategies can influence humans' feelings towards the group. For instance, Luria et al. [34] found that people were comfortable with re-embodiment if interactions were smooth and efficient, but when the robots required individual areas of expertise, such non-humanlike behavior raised concerns. Similarly, Reig et al. [37] found that re-embodiment was only accepted in contexts where personalization was expected, and that it was otherwise found as creepy or untrustworthy. Further, research by Williams et al. [11] suggests that the number of bodies and identities involved in a user's mental model dictates where and how they believe trust can be placed, and how they allocate trust to different *trust loci*. Trust in robots is essential in allowing robots to successfully take part in interactions and have people willingly interact with them [38]. In this work, we thus build off Williams et al. [11]'s notion of trust loci and seek to understand how identity performance strategies influence the trust placed and lost in the bodies and identities presented in a robot group.

Overall, previous research highlights the importance of the distinction between identity performance strategies, to understand both when it is appropriate for robot groups to present identity in certain ways, and how identity can be leveraged to design human-robot group interactions. The work described above suggests that robot group identity performance is critical because it may lead to differences in how groups of robots are mentally modeled, which may lead to differences in the perceived group entitativity, and the way trust is allocated to group members. Yet, the precise nature of the relationships between identity performance, entitativity, and trust remains unknown and untested. It is especially unclear if the relationship between identity performance strategies and trust is mediated by perceived entitativity. Thus, in this work, we seek to understand these relationships, by answering four key research questions:

RQ1: *How does identity performance strategy influence the perceived entitativity of a robot group?*

RQ2: *How does identity performance strategy influence how trust is distributed amongst the bodies and identities of a robot group?*

RQ3: *Does the perceived entitativity of a robot group mediate the relationship between identity performance strategies and trust distribution?*

RQ4: *After a blameworthy action, how does identity performance strategy influence how trust* loss *is distributed amongst the bodies and identities of a robot group?*

IDENTITY PERFORMANCE STRATEGIES AND THEIR HYPOTHESIZED EFFECTS

To begin to answer these questions, we consider the Robot Group Identity Performance Strategies described by Bejarano et al. [7], each of which is comprised of distinct Name and Voice Distinctiveness group identity observable cues.

All Unique: in which each robot body uses a unique name and a unique voice, conveying a unique identity.

All Shared: in which all robot bodies use the same name and the same voice, conveying a shared identity.

Only One: in which only one robot body speaks and uses a name, leaving ambiguous the nature of the other bodies.

For these strategies, we make the following hypotheses informed by previous work [7], [11]. For our first research question, we would hypothesize that:

H1: The order of perceived entitativity for the Identity Performance Strategies from least entitative to most entitative will be: the *All Unique* strategy followed by the *Only One* strategy followed by the *All Shared* strategy.

For our second research question, we would hypothesize that:

H2: *All Unique* strategy will lead to differing levels of trust in different identity-body pairings.

H3: *All Shared* strategy will lead to an even distribution of trust amongst the bodies and identities of the group.

H4: *Only One* strategy will lead to a greater level of trust placed in a named robot body and its associated identity than in unnamed, voiceless robot bodies.

For our third research question, we would hypothesize that:

H5: Perceived group entitativity will mediate the relationship between identity performance strategies and trust distribution.

For our fourth research question, we would hypothesize that:

H6: The *All Unique* strategy will lead to differing levels of trust lost in different identity-body pairings.

H7: *All Shared* strategy will lead to an even distribution of trust loss amongst the bodies and identities of the group.

H8: *Only One* strategy will lead to a greater level of trust lost in a named robot body and its associated identity than in unnamed, voiceless robot bodies.

METHODOLOGY

Experimental Design

To test these hypotheses, we conducted an IRB-approved online human-subjects study on the Prolific survey platform (prolific.co). Participants viewed two videos of a robot group interacting with a human in a simple airport help desk scenario. The study followed a 1x3 between-subjects design, with Robot Group Identity Performance Strategy as the independent variable. The three Robot Group Identity Performance Strategies described in the previous section were used, manipulated through the aforementioned changes in Name and Voice Distinctiveness observable cues.

Study videos contained subtitles to clarify which robot body was speaking (the color of the text matched the robot body color indicators) and the names used. Robots gestured using their head and hands whenever they were speaking, to make the communicating body visually obvious. All study videos, data, and analysis scripts are available via the Open Science Framework at https://bit.ly/3zeSbyO

Measures

The following Measures were used to assess the effects of our experimental manipulations on our dependent variables. To measure entitativity, participants were asked questions derived from previous robot group dynamics research [39], [26], [19]. Specifically, participants were asked to respond to the following prompts on 7-point Likert scales that best described their feelings or impressions of the video shown: (1) Do you think of the robots more as a group or more as unique, distinct individuals?, (2) The robots should be thought of as a whole, (3) How similar are the robots to each other?, and (4) How cohesive are the robots?.

To measure trust, participants were asked to complete the Reliability and Capability subscales of the Multi-Dimensional Measure of Trust (MDMT) Survey [40]. These subscales were presented four to six times: once for each of the three bodies and once for each identity (*All Unique* condition presented three distinct identities; *All Shared* and *Only One* conditions presented one identity). For each body, questions were prefaced by instructions to provide responses that best described their feelings or impressions of "the ROBOT with the [designated color] INDICATORS" and an image of the group with a specific robot body annotated. For each identity, questions were prefaced by instructions to provide responses that best described their feelings or impressions of an identity referenced by name. Both entitativity and trust questionnaires were followed by opportunities to provide brief open-ended explanations of responses.

Procedure

After providing informed consent and demographic information, participants were shown the first video of the study. Before the video, participants were told: "In this video, a human named 'Jane' interacts with 3 robots at an airport help desk." The participant was then shown a video with three Nao robots, each with different color indicators (red, blue, and green eye lights, bracelets, and chest tags) to distinguish the robot bodies from each other, as shown in Fig. 1. In the video, a human approaches and greets the help desk. Then, the robots introduce themselves according to the assigned condition: in *All Unique* each robot introduces itself separately with a unique name and voice, in *All Shared* all robots introduce themselves together with a shared name and voice, and in *Only One* a single robot introduces itself separately with a unique name and voice. The human then introduces themselves and says, "I am traveling to Denver, Colorado and need my boarding pass." The robots then ask for an identification card, process the request, and provide Jane with a boarding pass. After the video, participants completed the questionnaires described in the previous section.

Participants were then shown a second video. This video continued the interaction from the first video, and contained a blameworthy action. The presence of a blameworthy action in this second video allowed participants to more carefully consider where to place trust [41], and allowed us to measure not only how trust is placed, but also lost. Before the video, participants were told: "In this video, the human named "Jane" returns to the airport help desk and interacts again with the 3 robots." The participant was then shown a video with the same three Nao robots. In the video, the human approaches and greets the help desk again then indicates, "I just checked my boarding pass and the destination is incorrect" to which the robots ask for the incorrect boarding pass, process the request, and provide Jane with a new boarding pass. After the video, participants again completed the questionnaires.

Overall, the study lasted about 10 minutes. After completing the experiment, participants were compensated $2.

Participants

94 participants were recruited from Prolific (45 female, 45 male, 1 gender fluid, 1 non-binary, 2 N/A). Participants ranged from 19 to 68 years old (M=38.0, SD=13.2). 31 participants were assigned to *All Unique* condition; 32 participants were assigned to *All Shared* condition; and 31 participants were assigned to *Only One* condition.

Analysis

Composite Score Calculation: Pre- and post-test entitativity scores were calculated by averaging entitativity question responses. Pre- and post-test trust scores were calculated by averaging responses to trust questions for which participants did not select "does not apply". Trust loss scores were calculated by subtracting post-test from pre-test trust. Trust/trust loss distribution scores were calculated by averaging trust/trust loss score differences between pairwise combinations of bodies and identities:

$$\frac{\sum_{(e_1,e_2)\in B\cup I} |Trust(e_1) - Trust(e_2)|}{\#(e_1,e_2) \in B\cup I}$$

Here, B is the set of Bodies present, I is the set of Identities present, and each (e_1, e_2) is a unique symmetric pair in the union of these two sets, where symmetry implies that $(e_1, e_2) == (e_2, e_1)$.

Statistical Tests: A Bayesian statistical analysis was conducted on anonymized data using the JASP statistical software [42][1]. This analysis was comprised of a set of Repeated-Measures (RM) ANOVAs with Bayes Factor (BF) Analyses. Specifically, Inclusion BFs across Matched Models [44], [45] were calculated through Bayesian Model Averaging. The Inclusion BFs produced by this approach represent the strength of evidence in favor of models including each candidate main effect or interaction effect (relative to models not containing those effects) All BFs reported for RM-ANOVAs are thus $BF_{Incl_{10}}$, i.e., Inclusion BFs representing the odds ratio of evidence in favor of an effect (H_1) versus evidence against an effect (H_0). For all the following analyses, when sufficient evidence for an effect or difference was found (BF>3.0), the results were further analyzed using post-hoc Bayesian t-tests. The BFs reported for these post-hoc tests are again of the form BF_{10}, i.e., the ratio of evidence for an effect versus evidence against an effect.

To address RQ1, a Bayesian RM-ANOVA was used to analyze the differences in entitativity (dependent variable) between conditions (independent variable). The BF produced by this analysis indicates the strength of evidence in favor of a difference in entitativity between conditions.

To address RQ2 and RQ3, a Bayesian RM-ANOVA was used to analyze the differences in trust distribution (dependent variable) among the bodies and identities present in each condition (independent variable). In this analysis, entitativity was included as a covariate, to determine its effect on the relationship between trust distribution and condition. The BFs produced by this analysis indicate the strength of evidence in favor of a difference in trust distribution between conditions and the strength of evidence in favor of an effect by entitativity on that difference. Further, the data was split by condition and RM-ANOVAs with BF Analysis were used

[1]We used JASP version 0.16.3 due to the recent software changes that may affect the results for Bayesian RM-ANOVAs [43]

to assess specifically how trust (independent variable) is distributed among the different robot bodies and identities for each condition. The BF produced by this analysis indicates the strength of evidence in favor of a difference in the trust placed in the loci.

To address RQ4, a Bayesian ANOVA was used to analyze pre-test to post-test trust loss distribution among the bodies and identities present in each condition. The BF produced by this analysis indicates the strength of evidence in favor of a difference in trust loss distribution between conditions. Further, the data was split by condition and RM-ANOVAs with BF Analysis were used to assess specifically how trust loss (independent variable) is distributed among the different robot bodies and identities for each condition. The BF produced by this analysis indicates the strength of evidence in favor of a difference in the trust lost in the loci.

RESULTS

In this section, we present the results of our study, organized according to our posed research questions.

RQ1: How does identity performance strategy influence the perceived entitativity of a robot group?

The first analysis produced extreme evidence in favor of an effect of identity performance strategy on entitativity (BF=1.434×10^6). Post-hoc Bayesian t-tests indicated extreme evidence for differences in entitativity between the *All Unique* and *Only One* conditions (BF=1.535×10^6) and between the *All Shared* and *Only One* conditions (BF=4.115×10^7). Robot groups using the *All Unique* (pre-test M=75.468 SD=12.301; post-test M=74.355 SD=16.003) and *All Shared* (pre-test M=81.406 SD=11.995; post-test M=74.969 SD=20.009) strategies were perceived to have higher entitativity than robot groups using the *Only One* strategy (pre-test M=58.798 SD=17.317; post-test M=54.516 SD=18.960). These results do not support H1, as the order of perceived entitativity for the Robot Group Identity Performance Strategies from least entitative to most entitative was: *Only One* < *All Unique* < *All Shared*. Additionally, post-hoc tests indicated that there were no significant differences in entitativity between *All Unique* and *All Shared* (BF=0.360).

RQ2: How does identity performance strategy influence how trust is distributed amongst the bodies and identities of a robot group?

The second analysis produced extreme evidence for an effect of identity performance strategy on trust distribution (BF=1.291×10^6). Post-hoc Bayesian t-tests indicated extreme evidence for differences in trust distribution between the *All Unique* and *Only One* conditions (BF=3.237×10^7), and between the *All Shared* and *Only One* conditions (BF=1.044×10^{12}). Additionally, post-hoc tests indicated that there were no significant differences in the trust distribution of the *All Unique* and *All Shared* strategies (BF=0.665).

Robot groups using the *All Unique* (pre-test M=6.664 SD=7.844; post-test M=8.189 SD=14.856) and *All Shared* (pre-test M=5.832 SD=6.621; post-test M=3.616 SD=3.257) strategies had lower trust distribution scores (closer to 0) indicating that trust was more evenly distributed when these strategies were used. In contrast, robot groups using the *Only One* strategy (pre-test M=29.461 SD=20.806; post-test M=25.755 SD=18.852) had higher trust distribution scores, indicating greater differences in trust placed in the different available trust loci. After observing this effect, we split our data by condition and used additional RM-ANOVAS to verify these differences (or lack thereof) and to further assess specifically how trust is placed in the robot bodies and identities for each condition. We discuss these results separately for robot groups using different strategies.

Robot Groups Using the All Unique *Strategy:* This analysis provided anecdotal to very strong evidence against an effect of locus of trust (i.e., which body or identity was asked about) on all measures of trust ($0.017 \leq$BF≤ 0.369). These results thus provide evidence against H2: Robot groups using the *All Unique* strategy had trust evenly distributed across identities and bodies (Fig. 2).

Robot Groups Using the All Shared *Strategy:* This analysis provided anecdotal to strong evidence against an effect of locus of trust on all measures of trust ($0.048 \leq$BF≤ 0.350). These results thus provide evidence for H3 (Fig. 2).

Robot Groups Using the Only One *Strategy:* This analysis indicated extreme to strong evidence in favor of an effect of locus of trust on all measures of trust (pre-test BF=$2836.838 \leq$BF ≤ 7553.981; post-test BF=$20.634 \leq$BF≤ 27.722). Post-hoc Bayesian t-tests provided anecdotal to strong evidence in favor of differences in trust between the red and blue bodies ($1.992 \leq$BF≤ 26.378), red and green bodies ($1.773 \leq$BF≤ 18.203), red identity and blue body ($2.803 \leq$BF≤ 26.610), and red identity and green body ($2.717 \leq$BF≤ 17.367) for all measures of trust. In all cases, trust placed in the blue and green bodies was lower than the trust placed in the red body and identity as shown by Fig. 2. Meanwhile, the t-tests also provided anecdotal evidence in favor of to moderate evidence against differences in trust between the red body and red identity ($0.254 \leq$BF≤ 1.251) and anecdotal to moderate evidence against differences in trust between the blue and green bodies ($0.253 \leq$BF≤ 0.488) for all measures of trust. These results thus provide evidence for H4 (Fig. 2).

RQ3: Does the perceived entitativity of a robot group mediate the relationship between identity performance strategies and trust distribution?

In the second analysis discussed in the previous section, entitativity was set as a covariate, to determine its effect on the relationship between trust distribution (dependent variable) and identity performance strategy (independent variable). This analysis produced anecdotal evidence against an effect of entitativity on the trust distribution-strategy relationship (pre-test BF=0.477; post-test BF=0.862). These results provided evidence against H5.

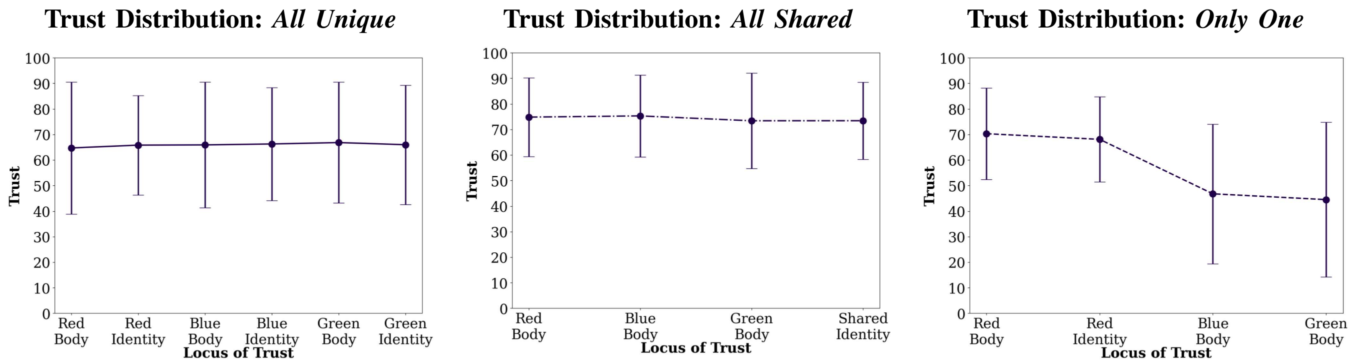

Fig. 2: Pre-test average trust placed in the robot bodies and identities. Error bars represent Standard Deviation.

RQ4: After a blameworthy action, how does identity performance strategy influence how trust loss *is distributed amongst the bodies and identities of a robot group?*

The final analysis produced anecdotal evidence against an effect of identity performance strategy on trust loss distribution (BF=0.694). All identity performance strategies had low trust loss distribution scores (*All Unique* M=10.736 SD=14.147; *All Shared* M=7.743 SD=7.336; *Only One* M=14.939 SD=15.639) indicating that trust loss was evenly distributed in all conditions. As before, we then split our data by condition and used RM-ANOVAS to verify this lack of differences in the trust lost in the robot bodies and identities for each condition. We discuss these results separately for robot groups using different strategies.

Robot Groups Using the All Unique *Strategy:* This analysis indicated very strong evidence against an effect of locus of trust on all measures of trust ($0.019 \leq BF \leq 0.026$). These results thus provide evidence against H6: Robot groups in the *All Unique* condition had the same level of trust lost in all identities and bodies.

Robot Groups Using the All Shared *Strategy:* This analysis indicated moderate to strong evidence against an effect of locus of trust on all measures of trust ($0.051 \leq BF \leq 0.166$). These results thus provide evidence for H7.

Robot Groups Using the Only One *Strategy:* This analysis provided anecdotal to moderate evidence against an effect of locus of trust on all measures of trust ($0.147 \leq BF \leq 0.485$). These results thus provide evidence against H8: Robot groups using the *Only One* strategy had the same level of trust lost in all identities and bodies.

Discussion

In this section, we discuss the implications of our results as guided by our research questions.

RQ1: How does identity performance strategy influence the perceived entitativity of a robot group?

Our results indicated that identity performance strategy has a significant effect on perceived entitativity. However, our results did not support our hypothesized order of perceived entitativity for the Robot Group Identity Performance Strategies from least entitative to most entitative (H1). Robot groups using the *All Unique* and *All Shared* strategies were found to have similar levels of perceived entitativity and higher levels of perceived entitativity than robot groups using the *Only One* strategy. These results differ from previous findings by Bejarano et al. [7], in which there were significant differences between all strategies that supported H1.

This deviation from previous research may be due to our use of only one speech observable to manipulate robot group presentation, whereas Bejarano et al. [7] manipulated multiple observables. Thus, human interactants may need multiple cues to perceive differences across different robot groups. Additionally, this could indicate that task ability and appearance may be better cues to prompt different perceptions of entitativity than identity presentation through speech. In their free responses, participants indicated that the robots all looked similar and acted in a similar manner. As one participant mentioned, “The robots seemed to be one solid unit. They worked together to complete the task.” Meanwhile, in the *Only One* condition, some participants' free responses indicated that they viewed all robots as looking similar and possibly having similar capabilities. However, as one participant indicated “it's not shown what the other two robots roles are so I'm not certain if they're very similar or dissimilar.” This suggests that the lower levels of perceived entitativity among this particular group may be due to the ambiguous nature of the unnamed, voiceless bodies.

RQ2: How does identity performance strategy influence how trust is distributed amongst the bodies and identities of a robot group?

Our results did not provide support for H2, as robot groups using the *All Unique* strategy had low trust distribution scores, with similar trust placed in each robot body and identity. In contrast, our results did provide support for H3, due to the similarly low trust distribution scores for robot groups using the *All Shared* strategy. Both of these results may be due to the higher perceived entitativity of the groups. Since all robots were seen as similar, the even trust distribution we observed may be due to a lack of reasons to trust one body/identity more or less than other bodies/identities. Thus, there may need to be greater signal of identity to indicate the uniqueness of identities presented, or uniqueness of the individuals within a group.

Our results also provided support for H4, as robot groups using the *Only One* strategy had higher trust distribution scores, with clear differences in trust placed in the robot bodies and identities within the group. Specifically, trust in the blue and green robot bodies was similar to each other, but was lower than the trust placed in the red body and its associated identity, demonstrating that a greater level of trust was placed in the named robot body and its identity. In the study videos, the blue and green robot bodies did not interact in any way, giving less of a signal as to where trust could and should be placed. Meanwhile, the red body presented an identity and was fully involved in the interaction, providing a social presence that the human could approach for help.

RQ3: Does the perceived entitativity of a robot group mediate the relationship between identity performance strategies and trust distribution?

Our results for RQ1 and RQ2 may hint at entitativity having a mediating role in determining the trust dynamics among groups using different identity performance strategies. However, our direct analysis of this relationship unexpectedly indicated that entitativity does not explain the relationship between group identity performance strategy and trust distribution. As such, this may indicate that the relationship between identity performance strategy, entitativity, and trust distribution is more complicated than expected.

RQ4: After a blameworthy action, how does identity performance strategy influence how trust loss *is distributed amongst the bodies and identities of a robot group?*

Our results provided support only for H7. Yet, we cannot confidently provide any conclusions surrounding this question, as regardless of identity performance strategy, there was evidence against any significant differences in trust lost in any robot bodies or identities (i.e. trust loss was evenly distributed). Nevertheless, these results raise interesting points about how human trust in a robot group is lost.

One might conclude that the even trust loss distribution for the *All Unique* and *All Shared* conditions was due to higher perceived entitativity, which might have rendered all bodies and identities as equally untrustworthy after the blameworthy action. However, this is at odds with our findings for RQ3, and with our observation that the *Only One* condition also had trust loss evenly distributed, but had a lower perceived entitativity and a greater level of overall trust placed in the named robot body and identity. These effects may have been observed because in all conditions blame could not be directly attributed to a specific robot body or identity; and since all robot bodies looked the same, the same level of trust was lost in all bodies and associated identities (if any).

Additionally, these findings may point to how trust (and trust lost) in the *whole group* influences trust in the *individual members* of that group. As previously mentioned, blame could not be directly attributed to a specific robot body or identity, so participants may have instead attributed the blamesworthy action to the overall group. If so, the loss in trust in the group as a whole may have manifested as similar levels of trust lost in each member of that group. Future work could explore the distinction between group and individual trust, and when this distinction is important, in the same way that Deconstructed Trustee Theory [11] explores the distinction between body and identity trust, and when that distinction is important. Understanding the relationship between group and individual trust may help us better understand both the antecedents and consequences for interaction quality of these different types of trust.

Limitations and Future Work

Although the robot group identity performance strategies considered in this work were presented through videos containing both speech and gestures, the group identity observable was only speech-based, with gesture only used to make the speaker more obvious, and to make the robots' communication seem more natural. Robot appearance and behavior obviously have a sizeable influence on human perceptions of robots [18], [46], [47], [48], [49], [50] and have long been used to personalize and give personality to individual robots. As such, the ways that visual cues may manipulate robot group identity should be further explored.

Additionally, while the experimental videos provided observers with a meaningfully interaction (simple airport help desk scenario) to reflect on while minimizing the influence of factors beyond those we manipulated, this simple interaction was of course not representative of actual human-robot interactions. As such, future experiments should include in-person human-robot interactions rather than observation studies alone. Such experiments will provide the opportunity to observe how participants would naturally interact with robot groups using different identity performance strategies and to interview participants on their perceptions.

Finally, future work should explore how trust distribution among trust subscales beyond reliability and capability are influenced by different identity performance strategies.

Conclusion

In this work, we explored how different identity performance strategies might differentially effect how users mentally model and allocate trust to robot groups. Overall, our work suggests that the presence (or lack) of identities within a robot group influences how humans distribute trust amongst loci within that group. Moreover, our results show that unnamed, voiceless robot bodies are seen as less trustworthy than bodies with associated identities; and the groups they are part of are seen as less entitative. These findings emphasize conclusions by Williams et al. [11]: understanding how different identity performances are perceived is critical to understanding how trust is built, retained, and lost. Future work towards this understanding should explore the role entitativity does play in the relationship between identity performance and trust dynamics, and how robot group identity performance strategies dictate how trust is gained and lost over more longitudinal interactions.

Acknowledgements

This work was supported by NASA Early Career Faculty award 80NSSC20K0070.

References

[1] S. Thiel, D. Häbe, and M. Block, "Co-operative robot teams in a hospital environment," in *Int'l Conf. intelligent computing and intelligent systems*, 2009.

[2] D. Claes and K. Tuyls, "Human robot-team interaction," in *Artificial Life and Intelligent Agents Symposium*, 2014.

[3] M. Shiomi, Y. Tamura, M. Kimoto, T. Iio, R. Akahane-Yamada, and K. Shimohara, "Two is better than one: verification of the effect of praise from two robots on pre-school children's learning time," *Advanced Robotics*, 2021.

[4] J. W. Thibaut and H. H. Kelley, *The social psychology of groups*, 1959.

[5] M. Sherif, "An outline of social psychology." 1948.

[6] C. H. Cooley and P. Rieff, *Social organization: A study of the larger mind*, 1909.

[7] A. Bejarano, S. Reig, P. Senapati, and T. Williams, "You had me at hello: The impact of robot group presentation strategies on mental model formation," in *Int'l Conf. HRI*, 2022.

[8] R. B. Jackson, A. Bejarano, K. Winkle, and T. Williams, "Design, performance, and perception of robot identity," in *HRI Workshop on Robo-Identity: Artificial identity and multi-embodiment*, 2021.

[9] A. M. Abrams and A. M. der Pütten, "I–c–e framework: Concepts for group dynamics research in human-robot interaction," *International Journal of Social Robotics*, 2020.

[10] P. A. Hancock, D. R. Billings, K. E. Schaefer, J. Y. Chen, E. J. De Visser, and R. Parasuraman, "A meta-analysis of factors affecting trust in human-robot interaction," *Human factors*, 2011.

[11] T. Williams, D. Ayers, C. Kaufman, J. Serrano, and S. Roy, "Deconstructed trustee theory: Disentangling trust in body and identity in multi-robot distributed systems," in *Int'l Conf. HRI*, 2021.

[12] M. R. Fraune, S. Sherrin, S. Sabanović, and E. R. Smith, "Rabble of robots effects: Number and type of robots modulates attitudes, emotions, and stereotypes," in *Int'l Conf. HRI*, 2015.

[13] M. Shiomi and N. Hagita, "Do the number of robots and the participant's gender influence conformity effect from multiple robots?" *Advanced Robotics*, 2019.

[14] H. Admoni, B. Hayes, D. Feil-Seifer, D. Ullman, and B. Scassellati, "Dancing with myself: The effect of majority group size on perceptions of majority and minority robot group members," in *Proc. Annual Meeting of the Cognitive Science Society*, 2013.

[15] G. Podevijn, R. O'grady, N. Mathews, A. Gilles, C. Fantini-Hauwel, and M. Dorigo, "Investigating the effect of increasing robot group sizes on the human psychophysiological state in the context of human–swarm interaction," *Swarm Intelligence*, 2016.

[16] M. Shiomi and N. Hagita, "Do synchronized multiple robots exert peer pressure?" in *Int'l Conf. human agent interaction*, 2016.

[17] E. Schneiders, E. Cheon, J. Kjeldskov, M. Rehm, and M. Skov, "Non-dyadic interaction: A literature review of 15 years of human-robot interaction conference publications," *J. Hum.-Robot Interact.*, 2022.

[18] S. Sebo, B. Stoll, B. Scassellati, and M. F. Jung, "Robots in groups and teams: a literature review," *Proc. ACM on HCI*, no. CSCW2, 2020.

[19] R. Wullenkord and F. Eyssel, "The influence of robot number on robot group perception—a call for action," *ACM THRI*, 2020.

[20] M. Dorigo, G. Theraulaz, and V. Trianni, "Swarm robotics: past, present, and future [point of view]," *Proc. IEEE*, 2021.

[21] A. Kolling, S. Nunnally, and M. Lewis, "Towards human control of robot swarms," in *Int'l Conf. HRI*, 2012.

[22] A. Fallatah, B. Stoddard, M. Burnett, and H. Knight, "Towards user-centric robot furniture arrangement," in *IEEE Int'l Conf. Robot & Human Interactive Communication (RO-MAN)*, 2021.

[23] D. F. Glas, T. Kanda, H. Ishiguro, and N. Hagita, "Simultaneous teleoperation of multiple social robots," in *Int'l Conf. HRI*, 2008.

[24] K. Zheng, D. F. Glas, T. Kanda, H. Ishiguro, and N. Hagita, "How many social robots can one operator control?" in *Int'l Conf. HRI*, 2011.

[25] M. R. Fraune, S. Kawakami, S. Sabanovic, P. R. S. De Silva, and M. Okada, "Three's company, or a crowd?: The effects of robot number and behavior on hri in japan and the usa." in *Robotics: Science and Sys.*, 2015.

[26] M. R. Fraune, S. Šabanović, E. R. Smith, Y. Nishiwaki, and M. Okada, "Threatening flocks and mindful snowflakes: How group entitativity affects perceptions of robots," in *Int'l Conf. HRI*, 2017.

[27] M. R. Fraune, B. C. Oisted, C. E. Sembrowski, K. A. Gates, M. M. Krupp, and S. Šabanović, "Effects of robot-human versus robot-robot behavior and entitativity on anthropomorphism and willingness to interact," *Computers in Human Behavior*, 2020.

[28] J. Berger, A. Bacula, and H. Knight, "Exploring communicatory gestures for simple multi-robot systems," in *Int'l Conf. Soc. Rob.*, 2021.

[29] A. Bacula and H. Knight, "Incoherent robot groups: How divergent motions within a robot group predict functional and social interpretations," in *Proc. Int'l Conf. Movement and Computing*, 2022.

[30] A. Nawroj, M. Toneva, H. Admoni, and B. Scassellati, "An exploration of social grouping in robots: Effects of behavioral mimicry, appearance, and eye gaze," in *Proc. Annual Meeting of the Cognitive Science Society*, 2014.

[31] H. Admoni, B. Hayes, D. Feil-Seifer, D. Ullman, and B. Scassellati, "Are you looking at me? perception of robot attention is mediated by gaze type and group size," in *Int'l Conf. HRI*, 2013.

[32] N. Salomons, M. Van Der Linden, S. S. Sebo, and B. Scassellati, "Humans conform to robots: Disambiguating trust, truth, and conformity," in *Int'l Conf. HRI*, 2018.

[33] K. Winkle, R. B. Jackson, A. Bejarano, and T. Williams, "On the flexibility of robot social identity performance: benefits, ethical risks and open research questions for hri," in *HRI Workshop on Robo-Identity*, 2021.

[34] M. Luria, S. Reig, X. Z. Tan, A. Steinfeld, J. Forlizzi, and J. Zimmerman, "Re-embodiment and co-embodiment: Exploration of social presence for robots and conversational agents," in *Designing Interactive Systems Conference*, 2019.

[35] S. Reig, E. J. Carter, T. Fong, J. Forlizzi, and A. Steinfeld, "Flailing, hailing, prevailing: perceptions of multi-robot failure recovery strategies," in *Int'l Conf. HRI*, 2021.

[36] S. Reig, E. J. Carter, T. Fong, A. Steinfeld, and J. Forlizzi, "Perceptions of explicitly vs. implicitly relayed commands between a robot and smart speaker," in *Int'l Conf. HRI*, 2022.

[37] S. Reig, M. Luria, E. Forberger, I. Won, A. Steinfeld, J. Forlizzi, and J. Zimmerman, "Social robots in service contexts: Exploring the rewards and risks of personalization and re-embodiment," in *Designing Interactive Systems Conference*, 2021.

[38] B. F. Malle and D. Ullman, "A multi-dimensional conception and measure of human-robot trust," *Trust in Human-Robot Interact.*, 2020.

[39] J. Spencer-Rodgers, M. Williams, D. Hamilton, K. Peng, and L. Wang, "Culture and group perception: Dispositional and stereotypic inferences about novel and national groups." *Jour. Pers. & Soc. Psy.*, 2007.

[40] D. Ullman and B. F. Malle, "Measuring gains and losses in human-robot trust: Evidence for differentiable components of trust," in *Int'l Conf. HRI*. IEEE, 2019.

[41] S. Guglielmo and B. F. Malle, "Asymmetric morality: Blame is more differentiated and more extreme than praise," *PloS one*, 2019.

[42] JASP Team, "JASP (Version 0.16.3)[Computer software]," 2022.

[43] D. van den Bergh, E.-J. Wagenmakers, and F. Aust, "Bayesian repeated-measures anova: An updated methodology implemented in jasp," Jun 2022. [Online]. Available: psyarxiv.com/fb8zn

[44] M. Hinne, Q. F. Gronau, D. van den Bergh, and E.-J. Wagenmakers, "A conceptual introduction to bayesian model averaging," *Advances in Methods and Practices in Psychological Science*, 2020.

[45] S. Mathôt, "Bayes like a baws: Interpreting bayesian repeated measures in jasp," *Cog. Sci. and more. Retrieved from: https://www. cogsci.nl/blog/interpreting-bayesian-repeated-measures-in-jasp*, 2017.

[46] A. Pipitone, A. Geraci, A. D'Amico, V. Seidita, and A. Chella, "Robot's inner speech effects on trust and anthropomorphic cues in human-robot cooperation," *arXiv preprint arXiv:2109.09388*, 2021.

[47] N. Dennler, C. Ruan, J. Hadiwijoyo, B. Chen, S. Nikolaidis, and M. Matarić, "Design metaphors for understanding user expectations of socially interactive robot embodiments," *ACM THRI*, 2022.

[48] L. Onnasch and C. L. Hildebrandt, "Impact of anthropomorphic robot design on trust and attention in industrial human-robot interaction," *ACM THRI*, 2021.

[49] M. Joosse, M. Lohse, N. V. Berkel, A. Sardar, and V. Evers, "Making appearances: How robots should approach people," *ACM THRI*, 2021.

[50] O. Herzog, N. Forchhammer, P. Kong, P. Maruhn, H. Cornet, and F. Frenkler, "The influence of robot designs on human compliance and emotion: A virtual reality study in the context of future public transport," *ACM THRI*, 2022.