Investigating Human Perceptions of Robot Capabilities in Remote Human-Robot Team Tasks based on First-Person Robot Video Feeds

Cody Canning, Thomas J. Donahue and Matthias Scheutz[1]

***Abstract*— It is well-known that a robot's appearance and its observable behavior can affect a human interactant's perceptions of the robot's capabilities and propensities in settings where humans and robots are co-located; for remote interactions the specific effects are less clear. Here, we use a remote interaction setting to investigate possible effects of *simulated* versus *real* first-person robot video feeds. The first experiment uses subject-level comparisons of the two video conditions in a multi-robot setting while the second and third experiments focus on a single robot, single video condition using a larger population (via Amazon Mechanical Turk) to study between-subjects effects. The latter experiments also probe the effects of robot appearance, video feed type, and stake humans have in the task. We observe a complex interplay between interaction, robot appearance, and video feed type as they affect perceived *collaboration*, *utility*, *competence*, and *warmth* of the robot.**

I. Introduction

There is converging evidence in the field of human-robot interaction (HRI) that a robot's *physical appearance and observable behaviors* can have important effects on human-robot interactions, particularly on the human's perception of the robot's capabilities, competence, and task performance [5], [11], [14]. More human-like robots have been shown to encourage emotional connection and trust of the human teammate in the robot [8], as well as perceived sociability [4]. Similarly, a robot's appearance and physical characteristics have been shown to drive the human's evaluation of that robot's trustworthiness [1], perceptions of competence and willingness to comply with instructions [5]. There is also evidence suggesting that the content and format of a robot's display play a similar role in modulating trust [2].

While the physical appearance and behavioral repertoire of robots can be observed in *local, co-located* human-robot interactions, *remote* human-robot interactions typically take place via screen-based interfaces where different aspects of the remote robot's state are displayed. Depending on the nature of the display, the human interactant might remotely observe the robot from a *third-person perspective* (e.g., if the robot is depicted on a map or an external camera shows the robot in the environment) or from a *first-person perspective* (e.g., if the robot's immediate surroundings are depicted as perceived by its laser-based or camera-based sensors).

Critical questions for HRI then are whether and to what extent (1) these screen-based interfaces influence human evaluations of remote robot teammates, (2) the *realism* of the interface can modulate human perceptions (e.g., real maps or video versus virtual maps or video), and (3) the human's knowledge of the robot's physical appearance and/or behavioral repertoire further modulate any interface-based effects.

Currently we have only preliminary, often-conflicting information with which to answer the above questions. For example, [10] showed that local interactions with a robot expressing urgency in its voice can lead to higher human-robot task performance when compared to remote interaction via cameras mounted in the environment. Another study, [11] reported many differences between local interactions with a robot versus those with a third-person remote interface showing a simulated robot. One difference was that at least in some cases subjects tended to issue more commands to the local robot than the remote, simulated one. Similarly [3] showed that humans enjoy interacting with and are more trusting of co-located robots than video-displayed robots. At the same time, [16] found that participants rated their comfort at being approached by a robot from different directions the same way in a video-based interaction as in a local interaction. In a similar study [7] found that an embodied co-located robot and the same robot displayed life-size on video produced roughly the same post-interaction ratings of dominance, trustworthiness, sociability, responsiveness, competence, and respectfulness.

These and other results suggest that in some cases the difference between first and third person interfaces for remote robots will affect human perceptions of those robots, while in other cases third person interfaces versus local interactions could yield the same results. Overall, it seems that several aspects, in addition to robot appearance and observable behaviors, might contribute to differences in human perception in local versus remote interactions, and with first-person versus third-person interfaces.

In this paper we set out to contribute to the resolution of these trade-offs by investigating the effects of first-person interfaces on human perceptions of robot capabilities, competence, and task performance. Specifically, we investigate whether the degree of realism – video from a real robot versus video from a simulated robot – can affect human performance in a mixed-initiative single-human multi-robot team task with robots of equal capability. Thus we rule out the possibility that any measured differences in task performance and participant ratings about the robots are the result of how capable the robots are. Further, we look into whether humans will have different beliefs about the robots based on their behaviors *without ever seeing any robot from a third-person perspective* (i.e., the humans never see what

[1]Cody Canning, Thomas J. Donahue and Matthias Scheutz are with the Department of Computer Science, Tufts University, Medford, MA, USA {cody.canning, thomas.donahue, matthias.scheutz}@tufts.edu

the robot looks like). In this case any results that differ between the two visual displays are likely the result of both the participants' preconceptions about robots in general and new beliefs formed about the robots in the task based on the videos. We begin with a description of the first experiment and our hypotheses for it, and then describe the surprising results that prompted two more experiments to probe for specific differences between first-person interfaces where subjects are either participants in or observers of the task.

II. Experiment 1

The first experiment was designed to study possible effects of realism of first-person interfaces on human performance and robot perception in a single-human multi-robot cooperative team task. Specifically, we wanted to investigate whether video feeds from a real camera mounted on a real robot compared to video feeds from a simulated camera on a simulated robot in a simulated environment would lead to differences in task performance and subjective reports of the robots as teammates. A secondary question was whether "robot autonomy" (i.e., the robot's ability to work toward completing the task goal autonomously if not commanded by the human) would be able to modulate the differences (as previous studies suggest it might [13]).

A. Materials and Methods

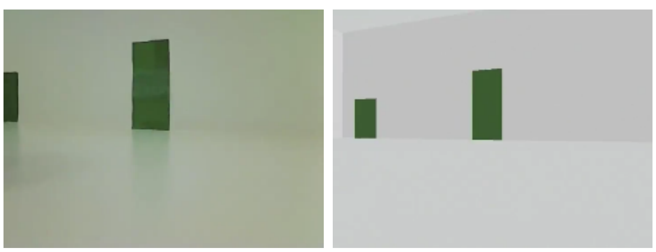

Fig. 2. Screen shots from real video feed (left) and simulated video feed (right).

Participants. 24 participants (12 female) recruited via a university website took part in this experiment. The average age was 20.88 (SD = 2.59) with a range from 18 to 31 years. All participants were right-handed and fluent English speakers with no history of brain trauma.

The Search and Report Task. Participants completed a HRI team task with a fictional storyline. At a computer the participant remotely supervised both a real robot and a simulated version of that robot in two different environments (known to participants as "red robot" and "blue robot", respectively). The participant's goal in the task was to explore with each robot simultaneously their respective environments to find a location in each environment where the field strength was strong enough to transmit important information back to "home base". Fields strengths varied from 0 to 1250 and a strength of 1200 or higher was required to succeed. Occasionally one of the robots (depending on task condition) would act autonomously, without the user's command.

Each robot's first-person view (via real or simulated camera) was depicted on screen. The blue robot's camera window was always below that of the red robot (as no vertical preference effects were expected). Each robot received commands (over a wireless connection for the real robot) from the participant as it drove around its environment and transmitted a live video feed. To the right of the camera views were windows containing clickable buttons which allowed the participant to issue the following commands to the respective robot:

- *Go Straight.* Robot drives straight.
- *Turn Left.* Robot turns left while maintaining forward velocity.
- *Turn Right.* Robot turns right while maintaining forward velocity.
- *Take a Reading.* Robot stops, assesses field strength for 2000 ms, reports field strength, continues with previous command (straight, right, or left).

Responses from the robots to commands and experiment notifications (such as time remaining, end of a trial, etc.). were displayed in popup dialogue windows to the right of the respective robot's camera view.

It is crucial to note that from the participants' perspectives the two robots were functionally identical; they were never told that one robot was simulated and the other real. Critically, participants neither saw nor heard the real robot at all, they only interacted with it remotely through the interface (similarly, they never saw any picture of the simulated robot). This was to ensure that they would not form any impressions or attitudes about the robot's capabilities based on visual appearance. Participants performed three six-minute trials, one in each of three conditions:

- *Real robot autonomy (RA):* the "red" robot (real) randomly entered into an autonomous subroutine interruptable by subject-issued commands.
- *No robot autonomy (NA):* neither robot was autonomous.
- *Simulated robot autonomy (SA):* the same as condition *RA* but for the "blue" (simulated) robot.

The autonomous subroutine loops on the following steps:

1) Stop for 2000 ms.
2) Move in the direction of highest field strength by going forward, turning left, or turning right for 5000 ms.

Participants could not succeed at the task as the transmission location was moved every two minutes and was always located outside the environment so that the robots could only get close enough to report a field strength of approximately 1150. This was done to ensure that every participant interacted with both of the robots for exactly six minutes per condition. We did not want to give participants the option to do the task with each robot serially.

Robots. The real robot was a Willow Garage `PR2` with a webcam attached to its base. The Agent Development Environment (ADE) implementation of DIARC [12] was the control architecture for both the real (`PR2`) and the simulated version of that robot.

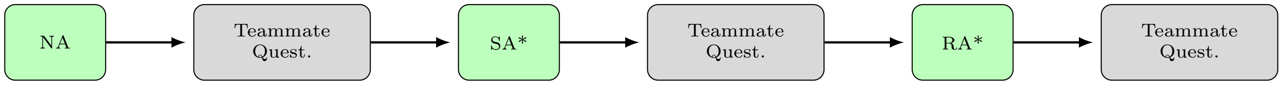

Fig. 1. Experiment 1 procedure. Asterisk denotes condition orders were counterbalanced.

Environment. The *real* robot operated in a 6.4 x 9.5 meter room with white walls, floor and ceiling. Landmarks and obstacles were placed in the room including a maroon box, a purple box, and two black boxes that each covered a door. Additionally, many colored rectangles (construction paper) were attached to the walls in order to provide visual landmarks in an otherwise overwhelmingly white room. A simulated version of the room was programmed to precisely replicate the physical one: color samples were captured with a webcam for reasonably accurate color reproduction and the room and landmark dimensions were the same.

Teammate Questionnaire. The robot's performance as a team-member in the search and report task was evaluated on a 9-point Likert scale with six items characterizing the robot (see Table I). Points 1 and 9 were labeled "Strongly Disagree" and "Strongly Agree", respectively.

Procedure. The participant first did a 5-minute practice run of the task in the *NA* condition. Once practiced, the participant completed all three conditions of the search and report task, followed by the teammate questionnaire for each robot (see Figure 1). These questions appeared in the same kind of popup windows as the robot responses and task notifications, to the right of the respective robot's camera view. Participants were always prompted about the red robot (top, real) first, and the blue robot (bottom, simulated) second.

B. Results

Given that the real and simulated environments are identical with respect to the dimensions and arrangement of obstacles, that the colors of obstacles and patches on the wall were very similar, and that the images of the real and simulated cameras were almost identical (only differing minimally with respect to shading and reflections, see Figure 2), we expected to see no differences in objective and subjective measures for equally capable robots (as the small differences in video quality should not affect the subjects' cognitive processes). We did, however, expect to see better objective performance and higher teammate ratings for the autonomous robots based on the results in [13].

Objective measures. Objective performance was measured by average distance from the transmission location. We characterized robots with a lower average as having performed the task better. A type-2 mixed-design analysis of variance (ANOVA) was conducted on the performance data with *between-groups* independent variables (IVs) Gender (male vs. female), Condition Order (NA, RA, SA vs. NA, SA, RA) and *within-groups* IVs Feed (real vs. simulated) and Robot Autonomy (yes vs. no). There were no significant main effects or interaction effects of any of these factors on objective task performance as expected for robots with equal capability, but unexpected for robots with autonomy compared to non-autonomous robots (see the discussion section below for a detailed explanation of the outcome).

Factor	1	2
This robot was helpful	.591	.665
This robot was capable	.419	.827
This robot was cooperative	.788	.500
This robot acted like a member of the team	.711	.551
This robot was easy to interact with	.782	.369
This robot was annoying	-.417	-.261
Eigenvalues	4.32	0.74
Cumulative Variance	0.41	0.72

TABLE I

OBLIQUELY ROTATED FACTOR LOADINGS FOR THE 6 TEAMMATE QUESTIONS. WHEN THE TWO GREATEST LOADINGS WERE INCLUDED (SHADED CELLS), THE ANALYSIS YIELDED A TWO-FACTOR SOLUTION.

Subjective measures. Exploratory maximum-likelihood factor analysis with varimax rotation was conducted on responses to the teammate questionnaire (see Table I). This analysis revealed two latent factors that accounted for 72% of the variance, p = .28 (indicating that we cannot reject the null hypothesis that the model described by the factors predicts the data well).

Collaboration: Two questions pertained to the participant's impressions of the robot as collaborative. The arithmetic means of responses from "this robot was cooperative", and "this robot was easy to interact with" were used to compose this new variable. Coefficient of internal consistency, Cronbach's α = 0.89, indicated that the scale was reliable.

Utility: Two questions pertained to the participant's impressions of the the robot's own task-related utility. The arithmetic means of responses from "this robot was helpful" and "this robot was capable" were used to compose this new variable. Cronbach's α = 0.89.

A type-2 mixed-design ANOVA was conducted on both latent variables with the same IVs. There was a significant main effect of Feed on *collaboration*, F(1,20) = 7.87, p = .01. Participants rated the real robot significantly higher (M = 4.76, SD = 1.78) than the simulated robot (M = 3.53, SD = 1.28). Similarly, there was a significant main effect of Feed on *utility*, F(1,20) = 11.56, p = .003. Participants rated the real robot significantly higher (M = 4.51, SD = 1.49) than the simulated robot (M = 3.38, SD = 1.78).

C. Discussion

That the real and simulated robots exhibited identical behavior in the task and that there were no *realism*-based effects on task performance together suggest that task performance

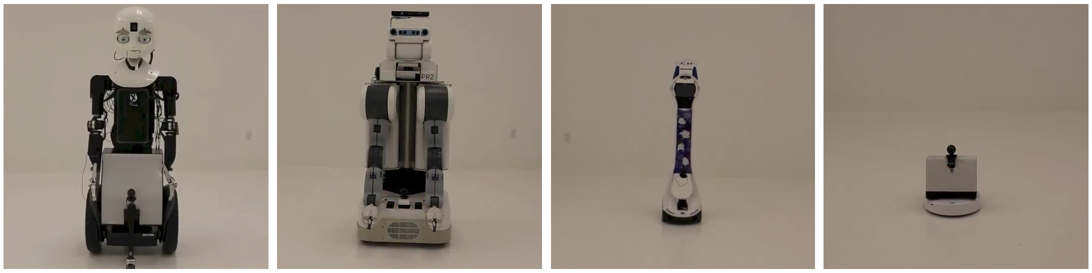

Fig. 3. Robots from the introduction videos arranged on a spectrum of human likeness. From left (most human-like) to right (least human-like): The Xitome Design `MDS`, Willow Garage `PR2` , VGo Communications, Inc. `VGo`, and iRobot `Create`.

of single-human multi-robot teams is not affected by the human's perception of *how the robot sees the world*, or related inferences about the nature of the robot. Perhaps surprising, however, are the subjective results, as they suggest subjects perceive robots as more collaborative and effective teammates when shown realistic first-person video feeds versus simulated (less realistic) ones.[1]

Moreover, it was surprising that some participants emphatically characterized robot autonomy as annoying and frustrating. In an effort to understand the causes, we found that the autonomy subroutine was too slow and too frequently interrupted to have the intended positive effects – participants switched between the two robots more rapidly than anticipated, causing the autonomous robot's behavior to regularly amount only to stopping. From the participant's perspective, this intermittent stopping was a hindrance rather than an assistance and can thus explain the unexpected lack of effects of robot autonomy on task performance.

III. EXPERIMENT 2

The unexpected differences in human perception of remote robots based solely on observations from first-person streaming video prompted a follow-up experiment to investigate possible biases toward real video compared to simulated video independent of partaking in the task. We employed a between-subjects paradigm wherein each participant saw the video feed from either the real or the simulated robot, but not both. This design precluded participants from (consciously or subconsciously) comparing the two video types and should reveal any preconceived notions about real robots or simulated ones (as they might be triggered by real versus simulated video feeds).

Moreover, by removing the interaction and multi-robot components and letting subjects evaluate a single robot's performance of the task with another human (without ever seeing the human or the robot), we can exclude possible influences on perceptions caused by the stake humans have in completing the task. Based on the results from Experiment 1, we hypothesized that participants would rate the robot with the real video feed higher on *collaboration* and *utility*, as well as perceived *competence*.

[1]It is also interesting to note that some subjects were shocked when, during debriefing, they learned that one of the robots was "real", suggesting that these preferences might be subconscious in at least some individuals.

Finally, to specifically explore the effects of prior knowledge about "robot appearance" and "observable behavior", subjects watched a brief introductory video of the robot approaching them (along with a no-video control condition). We hypothesized that seeing the introductory robot video would eliminate any preconceived differences between the simulated and the real video feeds. Since humanoid robots are often perceived as more intelligent than mechanical ones, and that taller robots are perceived to be more human-like than shorter ones [15], we also used robots with different human-like appearance and hypothesized that the more *human-like* the robot appears to be, the higher ratings of *collaboration*, *utility*, *humanness*, and *competence* the robot would receive.

A. Methods and Materials

The experiment was designed with the following between-subjects IVs:

1) Feed: factor with two levels, *sim* (video from simulated robot) and *real* (video from the real robot).
2) Introduction: factor with five levels, `NONE` (no robot introduction video), `MDS`, `PR2`, `VGO`, `CREATE`.

Participants. 137 individuals (48 female) participated in the experiment. Participants were Amazon Mechanical Turk workers that chose to complete the HIT (Human Intelligence Task) we posted. Participants were between 18 and 60 years of age (median = 31), residents of the United States, and native English speakers.

Search and Report Task Videos. The DIARC architecture was used again to operate both a physical `PR2` and a simulated robot for the filming of the search and report task videos. Each robot was equipped with a camera and executed the same series of action sequences while recording video. The camera in simulation was placed at the same height off the floor as a webcam attached to the base of the `PR2`. Responsiveness, acceleration and speed were kept constant between both robots. The same environment was used for these videos as in Experiment 1.

Videos for the *real* condition were recorded with the `PR2`'s camera and videos for the *sim* condition with the simulated robot's camera. Both robots followed the same preset routes through the environment. These videos were intended to depict *what the robot sees* during the task (from the 1st-person perspective) and at no point did they show the robot

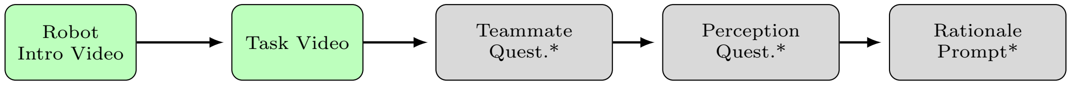

Fig. 4. Experiment 2 procedure. Asterisk denotes condition orders were counterbalanced.

itself. Three clips (between 25 and 40 seconds each) of the task were recorded in both the real environment and the simulated version. The three clips were edited together to form search and report task videos for each condition: in order to rule out potential ordering effects the three clips were permuted to form six task videos for each condition.

Video dimensions were 740 by 540 pixels, with a bitrate of 86 kbps. They were recorded at 30 frames per second. The videos depicted the robot navigating through the environment in response to commands issued by a human teammate. All dialog was subtitled and there was no audio. All human commands were presented in red and all robot responses in black. For all commands, if the robot was moving when the command was issued, it first stopped, then executed the command. The robot responded to all four directional commands (see Table II) with "OK", and to "Take a reading" with "Signal strength is *n*" where n is a value between 0 and 2000 based on the location of the robot relative to the transmission location.

Human Command	Robot Behavior
"Go straight"	Stops, then drives forward.
"Turn left"	Stops, then rotates counterclockwise.
"Turn right"	Stops, then rotates clockwise.
"Go back"	Stops, then drives in reverse.
"Take a reading"	Stops, then reports signal strength.

TABLE II

COMMANDS IN THE SEARCH AND REPORT TASK VIDEOS.

Human commands and robot responses were subtitled for 40 frames. Videos were edited to ensure *exactly* 45 frames were present between the onset of a command and the onset of the robot's corresponding behavior. If the robot had to stop it did so for *exactly* 25 frames before the onset of the corresponding behavior.

Robot Introduction Videos. The introductory robot videos were filmed for four different robots (see Figure 3). These robots are intended to form a spectrum of *human-likeness*. These videos were approx. 10 s and were composed of the following steps:

1) Robot stationary for 2 s.
2) Robot drives forward, toward the camera, at constant speed for 5 s.
3) Robot stops, facing the camera, for 3 s.

The arms of the `PR2` and `MDS` were positioned inertly at the waist of each robot and the face of the latter was configured to show a neutral expression. Each of the four robots was equipped with a webcam attached to base (approx. 1 foot from the floor). During the first 2 seconds of the "stopped" portion (item 3, above), the videos include a dashed red circle around the webcam with the label "camera" to indicate the camera through which the robot *sees* in the search and report task videos.

Perception Questionnaire. Three indices developed to gauge attitudes toward anthropomorphic characters were used to evaluate participants' perceptions of the robot from the *search and report* task [6], [9]. These indices assessed *perceived warmth, humanness* and *competence* on seven-point semantic differential scales of paired adjectives.

Rationale Prompt. Participants responded to the prompt: "Please describe briefly the reason(s) for your ratings on the previous pages". The previous pages are the perception and teammate questionnaires. This prompt was designed to elicit post hoc rationalizations.

Procedure. Participants took part in the experiment through a web survey hyperlinked from Amazon Mechanical Turk. After reading a short description of the study and consenting, participants supplied demographic information and were instructed about the steps of the task: part 1 consisted of watching a video of a human-robot search and report task (described in detail) that shows *what the robot sees during the task* and part 2 consisted of several questionnaires.

Participants (excepting those in the `NONE` condition) then proceeded to the introduction video page that read: *Before beginning Part 1, please watch this short video of the robot that will be participating in that task.* Following the introduction, participants watched the search and report task video and answered the questionnaires (see Figure 4). At the end of the experiment was a short "memory check" survey designed to filter out participants that did not pay attention to the videos.

B. Results and Discussion

Type-2 between-subjects ANOVA of the questionnaire data from experiment 2 revealed no significant effects of any of the IVs on any of the measures. Ratings of *collaboration* and *utility* were not significantly different across levels of Feed, Introduction or their interaction. That is, the effects seen in Experiment 1 were not present in this between-subjects design. Moreover, ratings of humanness, warmth, and competence were not significantly different by these factors. Hence, our original hypotheses were not confirmed by the outcome. As a result, we further hypothesized that the lack of differences between the video conditions must be due to the change in design to between-subjects from a repeated-measure design wherein participants were exposed to both the simulated and the real robot (and implicitly compared them) and/or the removal of the interaction of the subject with the robot. We also hypothesized that the lack of interaction might be the reason for the non-influence of the different introductory videos.

IV. Experiment 3

Experiment 3 extended Experiment 2 by adding in the interactive component from Experiment 1. We thus investigated an interactive protocol with a between-subjects design. Based on the previous results we expected that subjects interacting with the real robot would rate the robot higher on *collaboration*, *utility*, and *competence* than those interacting with the simulated robot. We further expected that priming with the introductory robot videos would eliminate these differences between the simulated and real video feeds (as hypothesized in Experiment 2), but *only* when interacting with the robot (not observing); without priming we might see the difference from Experiment 1.

A. Methods and Materials

Participants. 183 individuals (91 female) participated in the experiment. Participants were Amazon Mechanical Turk workers that chose to complete the HIT (Human Intelligence Task) we posted. Participants were between 18 and 60 years of age (median = 31), residents of the United States, and native English speakers.

Search and Report Interaction. A web interface was programmed for a "placebo version" of the interactive task where a video of the same type (*real* or *sim*) as the others for the same participant was displayed above a terminal window that produced notifications and communications from the robot. Positioned to the right of the video were five clickable buttons with the commands from Table II (replicating the interface from Experiment 1). The terminal produced a series of notifications (e.g., "initializing connection") to give the impression that the system was connecting to the real robot, after which a still of the first video frame appeared. Next, there was a five-second countdown to the start of the task, after which the video began to play.

Since the robot's camera feed was a pre-recorded video, the motion commands had no effect on the robot's behavior – however, there was always the chance that the robot's behavior would at times correspond to the most-recent command issued by the human. The only genuine interaction took place in response to the "take a reading" command, where a random field strength value (within 200 of the previous reported value) was printed in the terminal window. The rationale for this design was two-fold: (1) we surmised that the real interactions regarding readings would be sufficient for subjects to believe that the whole interaction was real; (2) pre-recorded videos removed the difficulties of running a large-scale study with real remote robots (which is technically hardly feasible); and (3) the occasional mismatch between a motion command and the robot's observed behavior would yield similar effects to the (failed) autonomy behavior in Experiment 1.

Procedure. The procedures of Experiment 3 are the same as those in Experiment 2 with the addition of the interactive search and report task followed by the same set of questionnaires and rationale prompt. This protocol allowed for the comparison of participants' appraisals after *observing* the robot and after *interacting* with it.

B. Results

A type-2 mixed-design ANOVA with between-subjects factors Feed, Introduction, and repeated-measures factor Task (*observe* vs. *interact*) revealed a significant effect of the interaction of Feed, Introduction and Task on ratings of *collaboration*, $F(4,163) = 2.86$, $p = .03$. Pairwise t-tests revealed significant differences between several pairs (see Table IV and Figure 5). Similarly, there was a significant effect of the interaction of Feed, Introduction, and Task on *competence* ratings, $F(1,163) = 2.46$, $p = .048$, and *warmth* ratings, $F(1,163) = 2.72$, $p = .03$. Pairwise t-tests revealed significant differences between several pairs for *competence* (see Table IV and Figure 5) and for *warmth* (see Table III).

There was a significant main effect of Task on *utility* ratings, $F(1,163) = 209.81$, $p < .001$. Participants rated

Feed/Intro	*M*, *SD*	Feed/Intro	*M*, *SD*	*p*-value
Real CREATE	.41, 1.06	Sim Create	-.09, .88	.053
Real CREATE	.41, 1.06	Sim NONE	-.10, .73	.054
Real CREATE	.41, 1.06	Sim PR2	-.46, .64	.001
Real CREATE	.41, 1.06	Real VGO	-.23, .87	.017
Real CREATE	.41, 1.06	Sim VGO	-.24, .89	.014
Sim CREATE	-.09, .88	Sim MDS	.56, .91	.011
Real MDS	.22, .93	Sim PR2	-.46, .64	.009
Sim MDS	.56, .91	Real NONE	.06, .78	.054
Sim MDS	.56, .91	Sim NONE	-.10, .73	.011
Sim MDS	.56, .91	Sim PR2	-.46, .64	.0001
Sim MDS	.56, .91	Real VGO	-.23, .87	.003
Sim MDS	.56, .91	Sim VGO	-.24, .89	.002
Real NONE	.06, .78	Sim PR2	-.46, .64	.047
Real PR2	.29, .80	Sim PR2	-.46, .64	.004
Real PR2	.29, .80	Real VGO	-.23, .87	.052
Real PR2	.29, .80	Sim VGO	-.24, .89	.044

TABLE III

Notable results from pairwise t-tests of *warmth* ratings post-interaction.

Feed/Intro	Feed/Intro	*p*-value
Sim MDS	Real MDS	.046
Sim MDS	Sim PR2	.002
Sim MDS	Real VGO	.001
Real PR2	Sim PR2	.104
Real CREATE	Sim CREATE	.122
Real CREATE	Sim PR2	.022

Feed/Intro	Feed/Intro	*p*-value
Sim MDS	Real MDS	.016
Sim MDS	Real NONE	.049
Sim MDS	Sim NONE	.040
Sim MDS	Sim PR2	.002
Sim MDS	Real VGO	.004
Sim MDS	Sim VGO	.020
Real PR2	Sim PR2	.152
Real CREATE	Sim PR2	.022

TABLE IV

Notable results from pairwise t-tests of *collaboration* (top) and *competence* (bottom) ratings post-interaction.

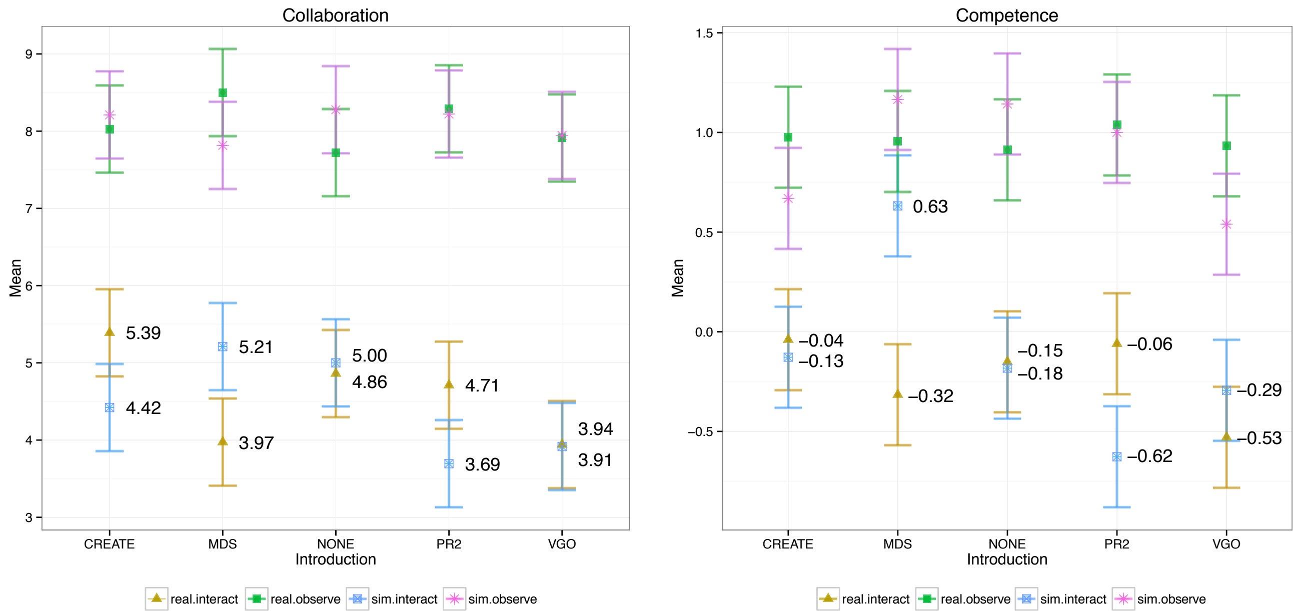

Fig. 5. Ratings of robot *collaboration* and *competence*. Means are provided for the *interact* condition. Error bars are Fisher's Least Significant Difference.

the robot in the *observe* condition (M = 6.84, SD = 1.72) significantly higher than the robot in the *interact* condition (M = 4.50, SD = 2.37). The interaction of Feed, Introduction and Task trended toward significance, F(1,163) = 1.96, p = .10, but was not significant.

We last observed a significant effect of the interaction of Feed and Task on *humanness* ratings, F(1,163) = 4.47, p = .04. Pairwise t-tests revealed significant differences between *real-observe* (M = -1.68, SD = 1.32) and *real-interact* (M = -2.06, SD = 1.06), p = .02, as well as between *real-observe* and *sim-interact* (M = -2.08, SD = 1.00), p = .02.

C. Discussion

We first reviewed the answers to the rationale prompts in order to gauge participants' beliefs about their level of interaction with the robot. Responses indicated that participants did in fact perceive the interaction to be genuine and believed they were working with a remotely located or simulated robot on the search and report task (there were *no* responses suggesting the participant thought the interaction was a prerecorded video, or that the task was "fake").

As expected, participants reported varying degrees of frustration with the robot based on their experience with it and the commands they issued (which the robot did not follow), suggesting that they had a stake in the task and a goal to find the transmission location as directed (this also confirms that the frustrating human-robot team task was successfully carried over from Experiment 1). Not surprisingly, the ratings of *competence*, *utility*, *collaboration*, and *warmth* were unanimously *lower* in the *interact* condition compared to the *observe* condition. Unexpectedly, results indicated that when *not* primed with the introduction video there were *no* differences in *collaboration*, *utility* and *competence* (as well as *warmth*) ratings between the real and simulated robots. This was the case in both the *observe* and the *interact* conditions. However, in some priming cases, ratings on these measures *did* differ.

On the *collaboration* and *warmth* measures the CREATE and the PR2 showed significantly (or trends toward) higher ratings of the real video feed. Similarly PR2 priming trended toward increased ratings of *competence* of the real robot over the simulated robot. Conversely, MDS priming increased ratings of both *competence* and *collaboration* of the simulated over the real robot. We interpreted this effect to be a result of the mismatch of high human likeness and ineffectiveness at the task. The *humanness* results showed that subjects who interacted with the robot (both *real* and *sim*) rated it lower than subjects who observed the *real* feed, suggesting that deficiencies exhibited by the robot during the interaction make it appear less human-like.

While the robot introduction primes did not eliminate differences between the real and simulated video feeds as predicted, they instead appear to cause differences between them. These data suggest that the knowledge of the robot teammate's appearance will affect perceptions of competence and collaboration (cooperation and ease of interaction), and, critically, that these affects vary not just in degree, but also in direction (e.g., inverse effects of MDS and PR2).

V. General Discussion

Experiment 1 showed that subjects completing a task with two robots characterized the robot with the real video feed as a more collaborative, higher utility teammate than the robot with the simulated feed, *without* having seen the robots.

Experiment 2 investigated using a larger population (via Amazon Mechanical Turk) whether people *not interacting* with the robot but instead *observing* a video feed from either the real or simulated robot of the same task in Experiment 1 would elicit a difference in teammate ratings or perceptions of humanness, warmth and competency of the robot. Results indicated that participants do not differ in their ratings, even if they are primed with a video that depicts how the robot appears and moves. Thus, we concluded that somewhere between observation of the single-robot task (Exp. 2) and authentic interaction in the two-robot task (Exp. 1), there is a catalyst for the results of Experiment 1.

Experiment 3 was designed as one way to probe for that catalyst: adding a "fake" interactive single-robot task that participants *believed* was real, while retaining the design of no subject-level comparisons of real and simulated feeds. Results suggested that for some robot primes there is a difference in ratings consistent with results from Experiment 1. Hence, the robot introduction prime, together with the participants' *belief that one is interacting* revealed a distinction between the simulated and real feeds (without having been able to compare them). It is thus likely that any possible comparison was not the *sole* cause for the results seen in Experiment 1. Experiment 3 showed that the belief that one is interacting with the robot on a task is required *in conjunction* with the robot's introduction prime for individuals to perceive a difference between the robot video feeds. In other words, the realism of the feed becomes important when the human teammate knows about the robot's appearance and they work together on a task. This is important because people's attitudes about robots' capabilities have been shown to influence their trust [1], [2], [5], which will affect their willingness to work with robots.

VI. Conclusions

In this paper, we presented results for systematic empirical investigations of the effects of *realism* in first-person robot video feeds from multiple remote robots in collaborative human-robot team tasks. The main results from the first experiment suggest that slight changes in the realism of video feeds might cause people to evaluate robots differentially *without* knowing anything about the robots' appearance or behavioral repertoire (other than what can be inferred from the interface). Two follow-up experiments suggested that when humans are not directly comparing video feeds of different realism and are only observing (not interacting with) the robot, these effects disappear, regardless of whether they know the robots' appearance or behavior. However, when humans believe that they are interacting with the robot, knowledge of robot appearance and behavior can bring about effects, modulated by the very robot appearance and behavior. The results thus suggest complex interplay between first and third-person displays of remote robots in collaborative team tasks and human knowledge of the robots' appearance and behaviors. These results are important for remote-interaction scenarios (e.g., a nighttime mission where the robot uses its sensors and night-vision camera to transmit video of the environment as perceived by machine vision, a representation that would filter out irrelevant objects and reduce noise to make the image meaningful to the human teammate).

In a next step, we plan to conduct a follow-up study wherein subjects will see the introductory robot videos in order to determine how the *realism* effect from Experiment 1 interacts with knowledge of robot appearance in a genuine interaction (with a spectrum of robot competence). We also intend to probe more deeply into what *about* the video feeds causes subjects to rate the robots differently by generating a spectrum of semi-real, semi-artificial looking video feeds that could provide more nuanced results.

References

[1] B. Adams, L. Bruyn, S. Houde, and P. Angelopoulos. Trust in automated systems: literature review. *Defence Research and Development Canada Toronto No. CR-2003-096*, 2003.

[2] H. Atoyan, J.-R. Duquet, and J.-M. Robert. Trust in new decision aid systems. In *Proceedings of the 18th International Conferenceof the Association Francophone d'Interaction Homme-Machine*, IHM '06, pages 115–122, New York, NY, USA, 2006. ACM.

[3] W. A. Bainbridge, J. Hart, E. S. Kim, and B. Scassellati. The effect of presence on human-robot interaction. In *Proceedings of Ro-Man*, pages 701–706. IEEE, 2008.

[4] E. Broadbent, V. Kuman, X. Li, J. Sollers, R. Q. Stafford, B. A. MacDonald, and D. M. Wegner. Robots with display screens: a robot with a more humanlike face display is perceived to have more mind and a better personality. *PLOS ONE*, 8(8):e72589, 2013.

[5] J. Goetz, S. Kiesler, and A. Powers. Matching robot appearance and behavior to tasks to improve human-robot cooperation. In *Proceedings of Ro-Man*, pages 55–60, 2003.

[6] C.-C. Ho and K. F. MacDorman. Revisiting the uncanny valley theory: Developing and validating an alternative to the godspeed indices. *Computers in Human Behavior*, 26(6):1508–1518, 2010.

[7] S. Kiesler, A. Powers, S. R. Fussell, and C. Torrey. Anthropomorphic interactions with a robot and robot-like agent. *Social Cognition*, 26(2):169, 2008.

[8] D. Li, P. P. Rau, and Y. Li. A cross-cultural study: Effect of robot appearance and task. *International Journal of Social Robotics*, 2(2):175–186, 2010.

[9] W. J. Mitchell, C.-C. Ho, H. Patel, and K. F. MacDorman. Does social desirability bias favor humans? explicit–implicit evaluations of synthesized speech support a new HCI model of impression management. *Computers in Human Behavior*, 27(1):402–412, 2011.

[10] R. Rose, M. Scheutz, and P. Schermerhorn. Towards a conceptual and methodological framework for determining robot believability. *Interaction Studies*, 11(2):314–335, 2010.

[11] P. Schermerhorn and M. Scheutz. Disentangling the effects of robot affect, embodiment, and autonomy on human team members in a mixed-initiative task. In *Proceedings of the 2011 International Conference on Advances in Computer-Human Interactions*, Gosier, Guadeloupe, France, 2011.

[12] P. W. Schermerhorn, J. F. Kramer, C. Middendorff, and M. Scheutz. DIARC: A testbed for natural human-robot interaction. In *AAAI*, 2006.

[13] E. Solovey, P. Schermerhorn, M. Scheutz, A. Sassaroli, S. Fantini, and R. Jacob. Brainput: enhancing interactive systems with streaming fnirs brain input. In *Proceedings of the 2012 ACM annual conference on Human Factors in Computing Systems*, pages 2193–2202. ACM, 2012.

[14] J. Wainer, D. J. Feil-Seifer, D. A. Shell, and M. J. Mataric. Embodiment and human-robot interaction: A task-based perspective. In *Processings of Ro-Man*, pages 872–877. IEEE, 2007.

[15] M. L. Walters, K. L. Koay, D. S. Syrdal, K. Dautenhahn, and R. Te Boekhorst. Preferences and perceptions of robot appearance and embodiment in human-robot interaction trials. *Proceedings of New Frontiers in Human-Robot Interaction*, 2009.

[16] S. N. Woods, M. L. Walters, K. L. Koay, and K. Dautenhahn. Methodological issues in HRI: A comparison of live and video-based methods in robot to human approach direction trials. In *Proceedings of Ro-Man*, pages 51–58. IEEE, 2006.