Engaging Robots: Easing Complex Human-Robot Teamwork using Backchanneling

Malte F. Jung[1], Jin Joo Lee[2], Nick DePalma[2], Sigurdur O. Adalgeirsson[2], Pamela J. Hinds[1], Cynthia Breazeal[2]

[1]Stanford University
Stanford, CA 94305 USA
{mjung, phinds}@stanford.edu

[2]Massachusetts Institute of Technology
Cambridge, MA 02139 USA
{jinjoo, ndepalma}@mit.edu, {siggi, cynthiab}@media.mit.edu

ABSTRACT
People are increasingly working with robots in teams and recent research has focused on how human-robot teams function, but little attention has yet been paid to the role of social signaling behavior in human-robot teams. In a controlled experiment, we examined the role of backchanneling and task complexity on team functioning and perceptions of the robots' engagement and competence. Based on results from 73 participants interacting with autonomous humanoid robots as part of a human-robot team (one participant, one confederate, and three robots), we found that when robots used backchanneling team functioning improved and the robots were seen as more engaged. Ironically, the robots using backchanneling were perceived as less competent than those that did not. Our results suggest that backchanneling plays an important role in human-robot teams and that the design and implementation of robots for human-robot teams may be more effective if backchanneling capability is provided.

Author Keywords
Affect; human-robot interaction; human-robot teams; team performance; urban search and rescue.

ACM Classification Keywords
I.2.9; K.4.3.

INTRODUCTION
Robots are part of high-stakes search and rescue operations [36], high-risk surgeries [25], and exploration missions. With ever-increasing capabilities, autonomous robots are becoming an integral part of professional and operational teams as people try to extend human capabilities. A burgeoning human-robot interaction research agenda is exploring the dynamics of human-robot teams with the goal of understanding more about how to enable coordination between people and robots [15, 44]. To date, this research has primarily focused on cognitive aspects of team functioning such as the development of situational awareness [37], common ground [48], task coordination [27, 44], and decision making.

CSCW '13, February 23–27, 2013, San Antonio, Texas, USA.

The role of subtle social signaling behavior in facilitating human robot teamwork, however, remains underexplored despite encouraging initial research highlighting the need for robots to be more adept at social interaction. By social signaling behaviors we mean behaviors that do not have their primary role in communicating content but rather in coordinating (e.g. when is it appropriate to speak?) and orienting (e.g. status and affect) participants of an interaction. By studying robot supported rescue teams in the field, the need for robots to have social interaction capabilities became evident when Fincannon et. al., [22] observed that rescue specialists engaged in social interactions with the robots even though the robots were not designed with social signaling capabilities. Bethel and colleagues [5, 6] proceeded to explore how robots without human-like social interfaces could exhibit socio-emotional behavior. Additional studies have demonstrated the powerful effects that simple social signaling behaviors can have in human-robot interaction. Only slight orientation towards a person, for example, can indicate attentiveness and caring [7, 11, 23]. The human-robot interaction community has so far primarily focused on how this signaling makes people think and feel *about the robot* itself [6], often with the goal of improving communication between the person and the robot. The primary goal of this research has been to make robots more understandable, intuitive, and predictable (or "believable") by using patterns that allow people to apply mental models and heuristics from interactions with people to infer a robot's internal states and intentions [9, 10, 24, 38, 42].

With the exception of two studies, the role of social signaling in improving *task-related outcomes* in human-machine interaction settings has been overlooked. In an experimental study, Breazeal and colleagues [8], for example, show that nonverbal social behaviors such as gaze, shifts in posture or orientation serve not only a cosmetic but also a pragmatic role in improving team functioning. Parasuraman and Miller [40] also demonstrate that adhering to social etiquette rules such as when and how to interrupt in an automated system can drastically improve overall human-machine system reliability.

The possibility that subtle, and seemingly task-unrelated, social signaling can have important implications for the performance of human robot teamwork is intriguing, and past research leaves two important questions unanswered.

First, it is not clear how these social signaling behaviors affect task outcomes that go beyond the interaction itself. For example do these behaviors have direct task-related benefits rather than only benefits in improving perceptions such as liking or naturalness? Second, it is not clear from past research whether the effects of social signaling behaviors extend beyond the simple dyadic human-robot scenarios to more complex human-robot team configurations.

A specific social signaling behavior that has been examined in the context of dyadic human robot interactions but not in complex human-robot teamwork settings is backchanneling [34, 46, 56]. Even though backchanneling is reported to have powerful effects in human-robot interaction, it is unclear whether those effects exist in complex scenarios in which the task, rather than the robot is the focus of attention. In the study presented here, we show that backchanneling is a subtle nonverbal social behavior that can have substantial effects on team functioning and perceptions of engagement in a complex teamwork scenario.

Backchanneling in Interactions

Backchanneling refers to a set of mostly nonverbal behaviors by a listener in a conversation, which signals to the speaker that the listener is actively engaged in the interaction. It includes behaviors such as "mm-hmm" vocalizations, slight nodding, eye contact, and orientation towards the speaker.

Within the CSCW community, backchanneling has largely been treated as form of concurrent feedback. Dennis and Kinney [19], for example, describe that backchanneling serves four cognitive functions including indicating understanding, indicating lack of understanding, repair or clarification of the message, and sentence completion. From this perspective backchanneling's main function is in establishing common ground by signaling that the receiver has understood the message [17]. Backchanneling, however, not only has a coordination function but also has a positive "gestalt emotional tone" [41]. It signals emotional attunement towards the speaker as an indication of emotionally positive validation, interest, and care about the partner's opinions [18].

Backchanneling In Human-Robot Teamwork

Given these broad cognitive and affective functions, backchanneling is a particularly promising behavior in the context of human robot interaction. First, it can be enacted nonverbally, which is important, as it can be manipulated without affecting the robot's verbal content. Second, it can be implemented mechanically and computationally less expensive by using simple head movements and reactive timings based on speaker cues. Third, its effectiveness in conveying responsiveness and smoothing coordination has been demonstrated in the context of human-robot interactions. Yamasaki et. al. [56], for example, showed that robot backchanneling in the form of simple turns towards the speaker was consistently reciprocated by human participants when being guided by a robot in a museum. Backchanneling, even when it is only embodied through head movement indicates engagement [45] and when expressed by an artificial agent over longer periods of time has been shown to communicate caring [7]. Despite previous research, however, the role of backchanneling in complex human-robot teamwork is, as of yet, underexplored.

Study Overview

To explore the implications of backchanneling for the functioning of teams, we designed a 2 (backchanneling present: yes vs. no) by 2 (complexity: low vs. high) experiment.

HYPOTHESES

Theories about human-human interaction have been shown to be transferrable not only to human-computer interactions [39] but also to human-robot interactions [49] and human-robot teams [48]. We therefore expect that the cognitive and emotional effects associated with backchanneling such as establishing common ground and signaling engagement can be leveraged to understand how people will behave in human-robot teams.

Backchanneling improves team functioning

We specifically anticipate that the benefits of backchanneling will aid team functioning in human-robot teams. Although subtle, the presence or absence of backchanneling indicated through nods, gaze, and "mhm" vocalizations has been shown to be a key factor in distinguishing high performance from low performance work teams [28]. Backchanneling has also shown to matter when people are interacting with robots. Bickmore and Picard [7], for example, showed that nonverbal behaviors such as head nodding when expressed by a computer agent strongly affected participants' willingness to work with the computer agent in a future task. Further, Wang and colleagues [52] found that a robot tracking and following a person's face increased enjoyment over a condition in which the robot remained still. In human-robot teams, we therefore predict that when robots use backchanneling, it will enhance team functioning. We further argue that this effect will be strongest when the task is complex. Complex tasks create more demand on team members, require more coordination as a result of greater uncertainty, and can generate more tension among team members than simple tasks [13]. We expect backchanneling through its cognitive role in establishing common ground and through its affective role in signaling positive attunement to alleviate the negative impact of increased complexity on a team.

H1: The presence of backchanneling by robots will improve team functioning (by decreasing stress and cognitive load and by increasing team coordination and performance) in human-robot teams, especially when the task is complex.

Backchanneling increases perceived engagement

Backchanneling can convey engagement in a task, suggesting that a team member is emotionally engaged and fully present [29]. This is true for robots as well. Sidner et al. [45], for example, showed that a robot capable of backchanneling behavior was perceived as more engaging to participants who were asked to jointly give a demo with a robot. We anticipate that this effect in human-robot interaction will transfer to human-robot team interaction. That is, we expect that when robots use backchanneling in a collaborative task, they will be seen as more engaged in the task.

We also argue that this effect will increase when the task is complex. A relationship between complexity and engagement was suggested by Carl Weick in his study of an airplane crew failure that led to a deadly collision [55]. He proposed that task complexity decreases as group engagement improves. That is, when people are fully engaged in a team task, they are better able to manage the complexity. We therefore anticipate that backchanneling will particularly help in complex tasks in which participants are trying to manage multiple activities and may need more overt signals from team members to help them cope with the complexity. In less complex tasks, however, backchanneling may be seen as less useful and therefore not an indicator of engagement.

H2: Backchanneling by robots will be seen as positively related to robots' engagement in the task when the task is complex, but less so when the task is not complex.

Backchanneling decreases perceived competence

Simple backchanneling used by robots might also have negative consequences. Human-like backchanneling behaviors can increase expectations team members have about the capabilities of a robot. Hinds, Roberts, and Jones [32], for example, showed that the more human-like a robot the more task responsibility people defer to it. If then expectations of the robot are not met, it might lead to disappointment and lowered perceptions of a robot's competence. Backchanneling can also directly indicate a lack of competence. Johnson [26], for example, conducted an experimental study and found that in a dyadic work interaction low authority employees, responsible only for mundane tasks, exhibited more backchanneling such as "mmm-hmm" vocalizations than their high authority counterparts who were responsible for complex decision making . That is, backchanneling was associated with lower status. This is in line with findings that associate backchanneling with decreased perceptions of reliability in robots [45]. Taken together, this research suggests that backchanneling can signal that the actor is lower status and less competent. We therefore anticipate that while a backchanneling robot may ease team functioning and be perceived as more engaged, it may also be seen as less competent and intelligent than a robot that does not exhibit any backchanneling behavior.

H3: Backchanneling by robots will decrease perceived competence and intelligence of the robot in human-robot teams, especially when the task is complex.

TASK DESIGN

To investigate our research questions, we required a task that utilized a human-robot team in which the people and robots were interdependent, required a high level of coordination and clear performance criteria, enabled us to vary task complexity, and created a sense of urgency for participants, thus fully engaging them in the task. Consistent with these requirements, we designed an urban search and retrieval (USAR) game in which human-robot teams worked together collaboratively to retrieve as many items as possible following a building collapse.

Urban Search and Retrieval Scenario

Based on available Urban Search and Rescue (USAR) scenario descriptions [36], we abstracted a USAR game described to participants as taking place in a large building that has recently collapsed, and in which the human-robot team needs to search and retrieve as many items as possible in 10 minutes. The USAR team consisted of the participant, two ground robots, and an aerial robot. The ground robots were mobile, upper-torso humanoid robots named Nexi and Maddox, and the aerial robot (UAV) was remotely controlled by a human operator (confederate) who was also described as a member of the team.

Figure 1. USAR Sencario with two humanoid robots (Nexi in front and Maddox in back)

Items (e.g. a briefcase, necklace, flashlight) were drawn onto small cards that were then covered by a post-it. Items could be found everywhere in the arena (see Figure 1) on the floor and also in baskets. Players needed to flip open the post-it to see what item was hidden underneath (see Figure 2). The goal of the game was to collect as many items as possible in 10 minutes. The team needed to coordinate together as each member has his/her own unique capabilities and skills in searching and retrieving the items that if utilized strategically could lead to better success.

The arena was split into three *danger zones* (low, medium, and high), mimicking the varying levels of safety and searchable access in USAR situations. A player could safely search in the low danger zone, but was not allowed in the high danger zone. To enter the medium danger zone, the player needed to check with the ground robots to see if access was safe. The ground robots were capable of searching in all danger zones, and if an item was in an area too dangerous for humans, the robot could retrieve that item for the player, when instructed to do so. The ground robots were also described as having a "special scanner" that allowed them to quick-scan a square (that it was currently in) to reveal all the items in that square. Acting as the "eyes in the sky" and flying in the high danger zone, the UAV+O (unmanned aerial vehicle and the operator) had better global situational awareness of where items were hidden and could give information about the specific location of an item to the participant. The confederate, acting as the UAV operator, responded with information from a script and had preplanned actions in controlling the flying robot.

A projector at the back of the arena displayed a set of three items to find for the current round and also a countdown clock indicating how much time remained. For example, Figure 2 displays a time of 3 minutes 18 seconds remaining and for the current round the player needs to find the items: footprint, vest, and map.

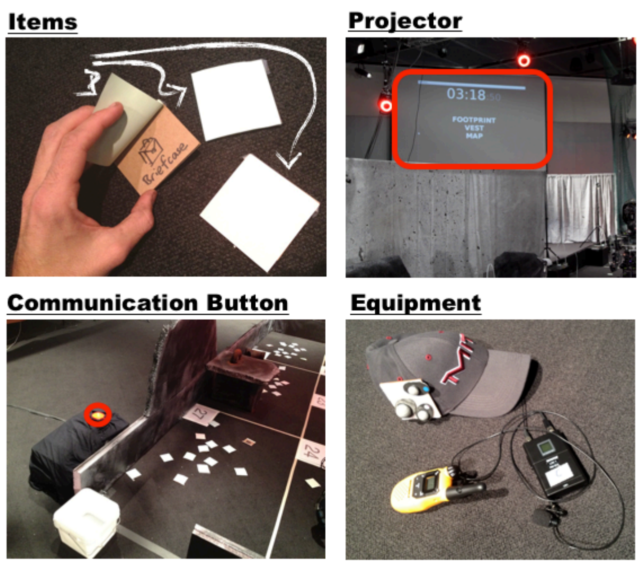

Figure 2. Photos of elements of the USAR game.

Game Play

The following narrative illustrates a typical game play from a player's point of view:

The game starts with a loud buzzer ringing in the arena, and I look up at the projector and see that time is counting down! Okay! We need to find the footprint, vest, and map for the first round. I get down on the floor and start looking in square 27 in the low danger zone. I frantically flip the post-its one by one to see what item is hidden underneath. I then remember that the ground robots have a special scanner and ask Nexi, "Can you scan square 25?" Nexi replies, "Okay I will go to square 25." Nexi and I both search in the low danger zone as Maddox goes off and searches around in the medium danger zone. I flip open a post-it and found the vest inside! I exclaim, "I have found the vest" into the mic so the robots can stop looking for that particular item. Nexi finally arrives in square 25 and after scanning the square says "I've found the footprint in square 25." I scramble over and start flipping and indeed found the footprint! Great, okay, just need to find the map to finish this round. I ask Maddox, "Have you found anything?" And he replies, "No, I have not found anything in particular yet." I grab my walkie-talkie and say, "Aerial operator, can you tell me where the map is?" He replies, "Okay, I'm on it!" I then see the UAV fly up into the air in the high danger zone as it surveys the entire scene. I continue looking around in the low danger zone, and I look over to see that both Nexi and Maddox are searching in the medium danger zone. The aerial operator finally reports back through the walkie-talkie, "The map is in the medium danger zone in square 8." I ask Maddox, "Is square 8 safe?" and he replies "No, it is not safe." I then ask Maddox to go into square 8 and once he arrives he says, "I have found the map in square 8, would you like for me to bring it to you?" I reply "Yes!" and moments later Maddox brings me a basket with the map inside of it and now its time for round 2!

TECHNOLOGY IMPLEMENTATION

Robot Platforms

The ground robots (Maddox and Nexi) are mobile, upper torso humanoid robots that are a part of the MDS (Mobile-Dexterous-Social) robot line as seen in Figure 1. The MDS robots are 4 to 5-feet tall with a hokuyo sensor mounted near the base of the robot and a kinect sensor mounted near the torso. They have 15 facial degrees of freedom (DoFs), 4 neck DoFs, a pair of 3 DoF shoulders, a pair of 5 DoF lower arms and hands, DoFs, a mobile wheel base with 2 DoFs, and a torso DoF. The neck and head mechanisms have 4 DoFs to support a lower bending at the base of the neck as well as pan-tilt-yaw of the head. The head can move at human-like speeds to support human head gestures such as nodding, shaking, and orienting.

The ground robots are fully autonomous agents (see below for system details), and the AR.Drone, a commercially available aerial robot, was teleoperated by a remote operator using its tablet interface.

System Overview

Autonomous Humanoid Robots

The role of the ground robots is to collaborate with the co-located human participant to successfully complete the USAR task. As such, these robots were capable of listening and recognizing human requests, verbally and nonverbally responding to humans, searching and navigating the arena, being aware of the task and environment states, and picking up baskets with items. To ensure consistent behavior of the robots, we developed fully autonomous robotic agents rather than utilizing a Wizard of Oz approach.

System Modules

By combining commercially available software solutions, open-source software packages, and custom in-house code, we created the following necessary technical modules relevant to the experiment:

Navigation: Laser range values from the hokuyo are used to estimate the robot's current position and orientation in a pre-made map of the arena through the Carnegie Mellon Robot Navigation Toolkit (CARMEN)[14]. We then used A* to path-plan to target destinations while avoiding static and moving obstacles. We included dynamic re-planning A* when an obstacle is hit.

Manipulation: The RGB image from the kinect was sent to ARToolKit [1], a software library that tracks special markers in physical space, to locate baskets in the 2D image. Unfortunately, since the robot's grasping range is outside the min range of the kinect's depth sensor, we estimated the 3D location of the basket in real coordinates using the z-depth value provided by the hokuyo, which can sense at a shorter minimum distance. We then used CCD (Cyclic Coordinate Descent [53]), an inverse kinematics method, to hook onto the basket's handle.

Speech: To understand the human utterances, we used the CMU Sphinx [47], an open source toolkit for speech recognition along with a human monitoring the speech input and either filtering or correcting misinterpreted speech. The robots spoke using synthesized voice through Cereproc [16], a real-time text-to-speech commercial technology, while also generating corresponding readable mouth positions with its articulated jaw. Nexi, the smaller robot, was given a female voice, and Maddox a male voice to make them easily distinguishable.

Human Localization: We used the Vicon [51] motion capture system to track a specially marked hat worn by the participant (see figure 2) and localize the human's position in the low danger zone. And by fusing the Vicon and CARMEN virtual spaces, the robots could also localize the human's relative position in the entire arena even when they were outside the vicon space.

Backchanneling: Backchanneling was provided through articulation of the head, eyes and upper torso. The feedback module was implemented by processing speech volume of the participant and then triggering the backchanneling response whenever the speech volume increased beyond a certain threshold. (For more detail, please consult the Experimental Manipulations section.)

System Behavior

The robot's primary behavior system was structured within a finite state machine that drove the main interaction, dialogue, and the random coverage search of the arena. A pilot study was performed early in the study design stage to determine the phrases most likely uttered in the task scenario, which served as event triggers for many states. The behavior system was a mixed initiative design in which an action queue could be self-populated and also managed by the interaction state machine. When a question or request was made to the robot, the robot responded by looking into its own memory or by addressing the command by prioritizing it in the action queue. In general, a separate action queue was maintained for managing the dialogue.

Task Management

The robot's task system was based on a common interaction policy between groups that was determined during the pilot study. A number of subsystems interacted by queuing and prioritizing social actions such as speech and gesture. The task model works closely with the dialogue in that the human can explicitly queue certain action policies such as searching for specific objects or visiting certain regions for exploration. Constraints were hard coded into the system to limit the potential actions. For instance, the robot could not go searching for an item once it had been asked to retrieve an object and the object was in its hand.

METHOD

To test our hypotheses, we conducted a 2 (presence of backchanneling: no vs. yes) x 2 (task complexity: low vs. high) between-subject experiment (N = 73).

Participants

We recruited ninety-six participants from a university community who were randomly assigned to one of the four conditions. Of the ninety-six participants seventy-three completed the study successfully and were included in the analysis – the others were dropped due to various malfunctions of the robots. Participants ranged in age from 18 to 40 (M = 25.0, SD = 6.1).

Experimental Manipulations

We manipulated backchanneling by adding or omitting nonverbal look-at and nodding behavior. In the backchanneling condition the robots oriented their body and turned their heads and eyes towards the participant, which composed of the "look-at" behavior, and nodded slightly whenever the participant made an utterance. We involved the entire "body" of the robots in the enactment of this behavior as the social perception of physical behavior is dependent on the entire body, rather than just the face [2]. We carefully timed the behavior so that it occurred in parallel with the utterances of a speaker and we chose a subtle, quick nod that was designed by a professional animator. The timing was to ensure that the looking-at and nodding would indicate concurrent backchanneling and listening feedback rather than an explicit "yes" response by the robot to something said by the participant. In the no backchanneling conditions, the robots did not exhibit any look-at or nodding behavior. As shown in Figure 3 below, requests by the human were always acknowledged verbally across conditions. Backchanneling was only manipulated through the presence and absence of nonverbal behavior.

We manipulated task complexity by altering two characteristics previously identified as fundamental task complexity characteristics [13], including multiple paths to a desired end state and conflicting interdependence. The presence of these characteristics alone or in combination distinguishes non-complex from complex tasks [13]. In the low complexity condition, we told participants that they would get one point for each item found. In the high-complexity condition we told participants that they would get one point for each item found, however, if they found all three items of one set in the color blue their points would double, if they found all in green their points would triple, and if they found all in red their points would quadruple. Each item was available in all three colors across conditions. Operationally the only change between low and high complexity was whether the task description gave the color added significance. This change in the task description opened up "multiple pathways to the desired end-state." Participants could vary their strategy to balance speed and the value of the combination of items. Participants also had to make tradeoffs as to whether they wanted to find more items within the time available or find the matching colors, which satisfied the "presence of conflicting inter-dependence" condition for task complexity. The order and composition of item sets were identical across conditions.

Materials and Measures

We used a range of subjective and objective measures that were administered during and after the task to test our hypotheses. Our dependent measures include four items for team functioning, two for robot engagement, and two for perceptions of the robots' competence.

Manipulation Checks

As a manipulation check of the robots' backchanneling, we asked participants to rate the extent to which they thought the ground robots were looking at them on a seven-point Likert type scale between "strongly disagree" and "strongly agree."

Our second manipulation, task complexity, was measured using a two-item scale (α = .70) comprised of items taken from [20] "I found this to be a complex task" and [35] "The task required me to coordinate many different things at the same time". Participants rated the items on a seven point scale from "totally disagree" to "totally agree."

Team Functioning

To address the first hypothesis, we measured stress, cognitive load, team coordination, and team performance. We measured participants' level of stress with a three-item sub-scale (α = .72) of the Positive and Negative Affect Schedule (PANAS) [54] administered as part of the post-task survey. Participants were asked to report on a five-point scale ranging from "very slightly or not at all" to "extremely" with regard to how much they were distressed, upset, and irritable during the task.

To measure cognitive load, we took a direct, objective measure using a dual task paradigm [12, 50]. Participants were instructed to monitor and press a big red "communication button" as soon as possible when it started to blink (see figure 2). The blinking button, they were told, meant that other responders were trying to verify that they were okay and pushing the button would signal that the team was safe. The button blinked three times at 3, 5, and 9 minutes into the task, and we measured the reaction times

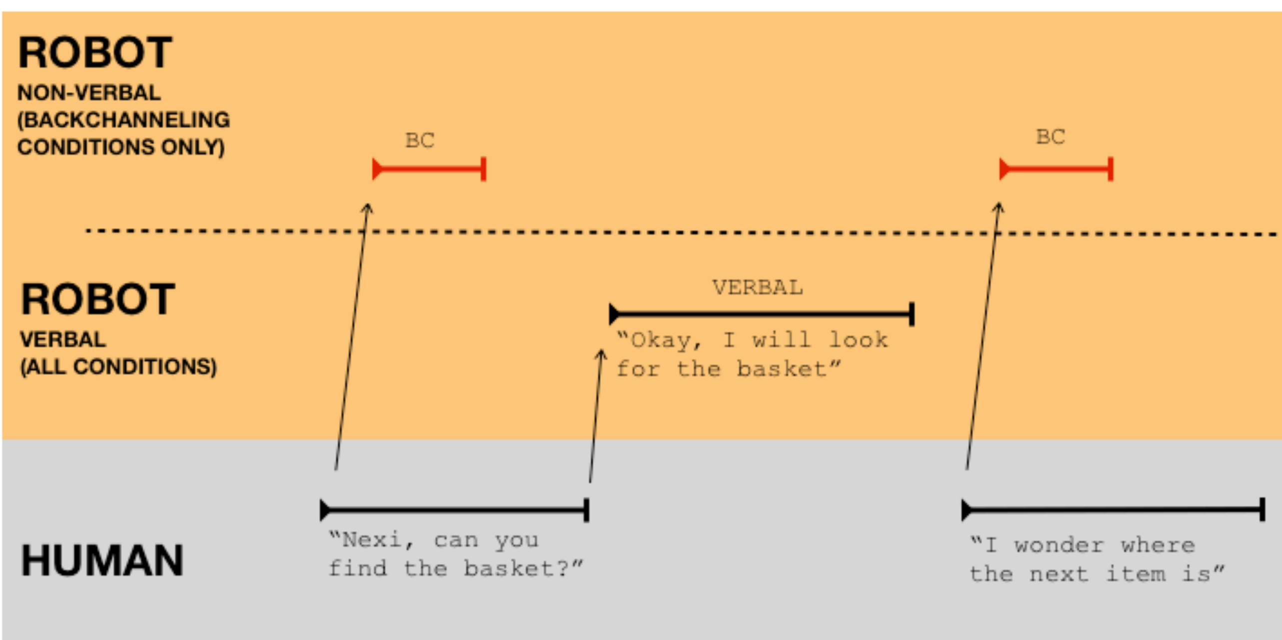

Figure 3. Temporal visualization of backchanneling manipulation. The robots respond verbally to requests across conditions, however the nonverbal backchanneling is only triggered in the backchanneling condition.

between the time the button started to blink and the moment the participant hit the button. We averaged across three measurements to account for the possibly varying travel distances by the participants. Our objective cognitive load measure was computed by taking the sum of all three reaction times and dividing the sum by the number of button presses that were made before the timeout. Several participants forgot about the button entirely and therefore reaction time was only used as a measure if the button was pressed at least once during the task ($N = 55$). We also used a self-report measure of cognitive load in which, as part of the post-task survey, we asked participants the extent to which "the task was mentally demanding" on a 7-point scale ranging from "totally disagree" to "totally agree."

We measured perceived team coordination with the five-item coordination subscale ($\alpha = .83$) from the Transactive Memory System (TMS) scale by Lewis and colleagues [33]. Participants rated items such as "Our team worked together in a well-coordinated fashion" on a 5-point Likert type scale ranging between "strongly disagree", and "strongly agree."

Finally, we measured team performance by recording the number of items found by each team. We counted how many correct items participants placed in the container within the allotted ten minutes of the task.

Perception of Robots' Engagement
To address our second hypothesis focusing on how backchanneling from the robots and task complexity would affect perceptions of the robots' engagement, we measured perceived distraction and responsiveness by using 7-point scale items ranging from "totally disagree" to "totally agree". Perceived distractedness used a 2-item measure including questions about the extent to which the ground robots seemed distracted and unfocused ($\alpha = .70$). We measured responsiveness by asking respondents about the extent to which they "felt that the ground robots were responsive."

Perception of Robots' Competence
To address our third hypothesis focusing on how robot backchanneling and task complexity would affect perceptions of the robots' competence, we employed two measures of robot perception in our post-task survey. First we measured perceived intelligence with a 5-point scale by asking "How much did you feel as if you were accompanied by intelligent beings?" borrowed from [32]. Second, we measured perceived competence by asking participants to rate the extent to which they "felt that the ground robots are very competent." on a 7-point scale.

Procedure
Upon arrival in the lab participants were asked to sit at a table separated from the task arena by a curtain and were asked for their informed consent to participate in our study. First, each participant filled out the pre-task section of our online survey, which asked about their demographic background, knowledge and experience of robots, and prior knowledge of the MDS robot platform that we used in our study.

Next, participants were introduced to the task through an instructional slide presentation – either the "low complexity" or "high complexity" version, depending on the condition to which they were assigned. Each participant was then asked to enter the task arena and was introduced to the physical setup of the space. An experimenter re-iterated the main points of the task and participants were given the opportunity to ask further clarification questions about the task. Before starting the main task, participants performed a trial search and retrieval task by finding and retrieving a specific item without the help of the robot. Participants were informed that the best performing team would win an additional $200. This task introduction and training procedure took approximately 15 minutes.

For the main task, participants conducted the ten-minute search and retrieval task as outlined in the previous section. Participants were asked to find and retrieve as many items as possible. After reaching the ten-minute time limit, we congratulated participants for their performance and asked them to return to the table behind the curtain to complete the post-task survey. Participants were then thanked and paid.

Data Analysis
We used a two-way analysis of variance (ANOVA) to investigate whether robot backchanneling and task complexity influenced team functioning and participants' perceptions of the robot. In each ANOVA we used robot backchanneling (no vs. yes) and task complexity (low vs. high) as independent variables. We also included participant gender, age, and prior knowledge about robots as covariates, but none of them had any significant influence on our dependent measures. We therefore omitted the covariates from further discussion.

RESULTS

Manipulation Checks
Participants in the backchanneling condition felt more strongly that the ground robots were looking at them ($M = 4.78$, $SE = 0.21$) as compared with those in the no-backchanneling condition ($M = 4.03$, $SE = 0.23$), $F(1, 65) = 5.93$, $p < .05$, providing evidence that our backchanneling manipulation was effective.

Interestingly, the main effect for perceived task complexity was not significant because the backchanneling manipulation interacted with perceived task complexity. We found a main effect for robot backchanneling on task complexity ($F(1, 69) = 4.78$, $p < .05$), as participants perceived the task as significantly less complex in the backchanneling condition ($M = 5.16$, $SD = 1.16$) than in the no-backchanneling condition ($M = 5.76$, $SD = 1.16$).

Team Functioning

In our first hypothesis, we argued that backchanneling by a robot would improve team functioning, especially when the task is complex. To test this, we examined participants' self-reported stress during the task, cognitive load as measured by reaction time, self-reported team coordination, and performance measured by number of items found in the task. H1 was partially supported.

Stress

As expected, we found a significant interaction between backchanneling and complexity of the task when predicting the amount of stress perceived by the participant $F(1,69) = 7.63$, $p < .01$. This indicates that when task complexity was low, the perceived stress of participants was similar in the no-backchanneling ($M = 1.72$, $SD = 0.70$) and backchanneling condition ($M = 1.87$, $SD = 0.50$). However, when task complexity was high, the perceived stress of participants was significantly higher in the no-backchanneling condition ($M = 2.55$, $SD = 0.86$) than in the backchanneling condition ($M = 1.78$, $SD = 0.73$). Additionally, there was a significant (positive) main effect of task complexity on perceived stress, $F(1, 69) = 5.00$, $p < .05$ (see left graph in Figure 4).

Cognitive Load

Consistent with hypothesis 1, we also found a significant main effect of robot backchanneling when predicting our behavioral measure of cognitive load $F(1, 51) = 6.70$, $p < .05$. The presence of backchanneling reduced the reaction time on the parallel task of pressing the button by an average 46% between the backchanneling condition ($M = 101.28$ sec, $SD = 62.40$) and the no-backchanneling condition ($M = 184.87$ sec, $SD = 157.21$), suggesting that people experienced significantly less cognitive load when the robot used backchanneling. Our self-report measure of cognitive load confirmed that the task was perceived as less mentally demanding when the robot used backchanneling, $F(1,69) = 4.91$, $p < .05$. There was, however, no interaction between backchanneling and task complexity, so hypothesis 1 was only partially supported (see right graph in Figure 4).

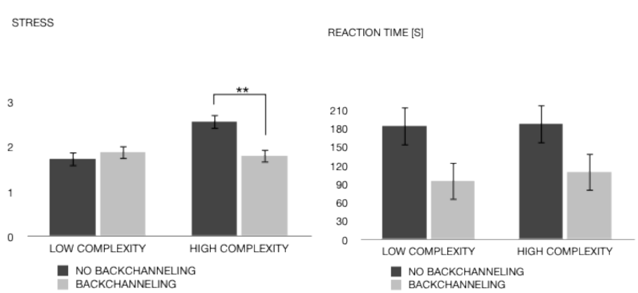

Figure 4. Left Graph: Effect of robots' backchanneling and task complexity on stress. Right graph: Effect of robots' backchanneling and task complexity on cognitive load (measured as response time in seconds). Error bars = ±1SE

Team Coordination

We found a significant main effect of task complexity on participants' sense of overall team coordination $F(1, 70) = 5.45$, $p < .05$. Participants found their teams to be more coordinated when task complexity was low ($M = 3.44$, $SD = 0.75$) in comparison to when task complexity was high ($M = 3.03$, $SD = 0.80$). There was not, however, any main effect of robots' backchanneling, nor an interaction effect between backchanneling and complexity, as predicted in hypothesis 1 (see left graph in Figure 5).

Team Performance

Our last indicator of team functioning was team performance measured by the number of items found during the task. Although the main effect of backchanneling on the number of items found was not significant, $F(1, 69) = 2.31$, $p = .13$, the direction was as predicted. That is, participants performed better, albeit not significantly, when the robot used backchanneling (see right graph in Figure 5).

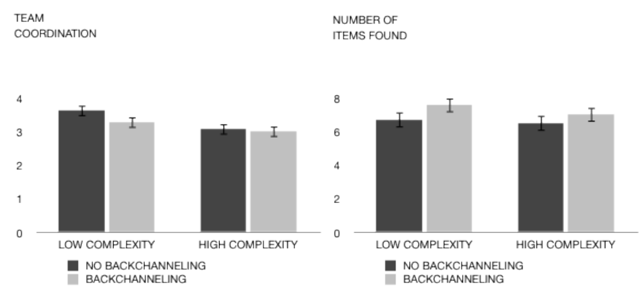

Figure 5. Left graph: Effect of robots' backchanneling and task complexity on perceived team coordination. Right graph: Effect of robots' backchanneling and task complexity on team performance (number of items found). Error bars = ±1S

Perceptions of the Robots' Engagement

We argued in hypothesis 2 that participants would perceive robots that used backchanneling to be more engaged and that this would be strongest in the high task complexity conditions.

Perceived responsiveness

Our analysis predicting perceived responsiveness supports hypothesis 2. The interaction between robots' backchanneling and task complexity was significant, $F(1,69) = 6.17$, $p < .05$, indicating that when the task was complex, the robots that gave feedback were seen as more responsive whereas in the less complex task, the robots that gave no backchanneling were seen as more responsive. There were no main effects (see left graph in Figure 6).

Perceived distractedness

Our analysis predicting distraction yielded a similar result. That is, when the task was complex, robots that used backchanneling were seen as less distracted whereas, in the low task complexity conditions, the robots that provided feedback were seen as more distracted, $F(1, 69) = 4.50$, $p < .05$. Thus H2 was partially supported (see right graph in Figure 6).

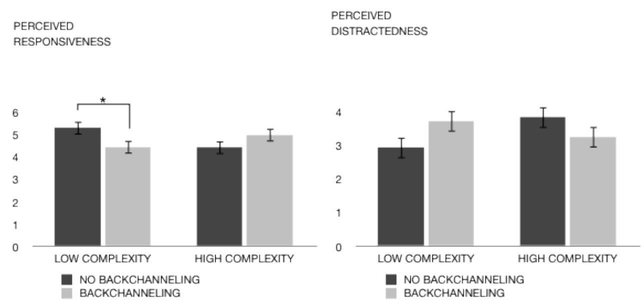

Figure 6 Left graph: Effect of robots' backchanneling and task complexity on perceived responsiveness of robots. Right graph: Effect of robots' backchanneling and task complexity on perceived distraction of robots. Error bars = ± 1 SE

Perceptions of the Robots' Competence

In hypothesis 3, we argued that participants, despite the positive effects of robots that use backchanneling, would perceive affective robots more negatively.

Perceived Intelligence

As expected, we found a significant main effect of robot backchanneling on perceived intelligence of the robots $F(1, 69) = 4.53$, $p < .05$. Participants found the robots that used backchanneling ($M = 2.68$, $SD = 1.02$) to be on average 16% less intelligent than the robots that did not use backchanneling ($M = 3.20$, $SD = 1.02$). We did not, however, find an interaction effect, as hypothesized (see left graph in Figure 7).

Perceived Competence

As with perceived intelligence, our analysis of perceived competence of the robot showed that robots that used backchanneling were perceived as less competent than robots that did not, $F(1,69) = 4.65$, $p < .04$. Again, there was no interaction effect, so hypothesis 3 was only partially supported (see right graph in Figure 7).

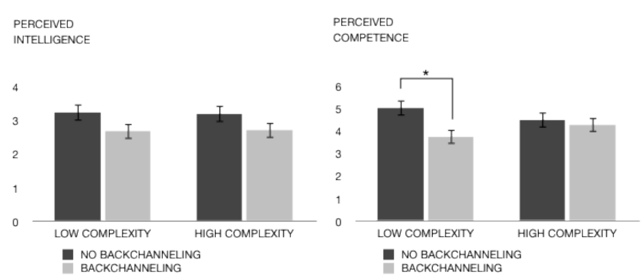

Figure 7 Left graph: Effect of robots' backchanneling and task complexity on perceived intelligence of robots. Right graph: Effect of robots' backchanneling and task complexity on perceived competence of robots.

DISCUSSION

We argued and found that subtle backchanneling by robots in human-robot teams helped team functioning (lower stress, lower cognitive load) and perceived engagement of the robots, especially when the task was complex, but at the same time lead to robots being seen as less competent. We also found that task complexity had a significant effect on peoples' responses to the robots' backchanneling. In general, we saw a pattern in which more benefits (less stress, less cognitive load, more perceived engagement) were found from backchanneling in high complexity conditions over low complexity conditions. Complexity, however, did not interact significantly with perceptions of robots' competence. These results suggest that the biggest benefits from backchanneling in human-robot teams may be seen when tasks are demanding and complex.

Although the trend was in the right direction, we were not able to demonstrate a significant effect of backchanneling on the number of items found. We attribute this to a ceiling effect. One of the items in the fourth round of the USAR game could only be picked up by a robot as the item was in the high danger zone at the far end of the arena. Retrieving this item took the robots an especially long time due to travel time. Even if all other items in that set had already been found, the participants waited for the robot to retrieve the item in the danger zone. All participants ran out of time before the robot retrieved that item and therefore no participant advanced from round 4. It would be useful if future research included more sensitive measures of performance to better detect the effects of backchanneling on performance.

Previous research has highlighted both a cognitive and an affective role of backchanneling in interactions. Cognitively, backchanneling is a key factor in establishing common ground [17] and affectively backchanneling indicates emotionally positive engagement and listening [18]. Contrary to our expectations, backchanneling did not have any effects on participants' perceptions of team coordination. This suggests that our particular way of embodying backchanneling was not successful in tapping into the cognitive benefits of backchanneling to build common ground. If our manipulation had worked primarily at a cognitive level, we would have expected the robots to be seen as more competent and intelligence, especially in high complexity conditions, because they would have been seen as instrumental in establishing common ground. We found, however, that backchanneling reduced perceptions of competence and intelligence independent of task complexity. Backchanneling might also have raised participants' expectations about the robots' cognitive capabilities. When these cognitive expectations were not met, people thought of the robots as less competent collaborators.

The results we obtained are more consistent with theories of backchanneling as an affective phenomenon. To understand backchanneling as affective, we take what Keltner and Haidt [30] called a social-functional perspective on affect. This perspective emphasizes the interpersonal functions of emotion and the notion that the emotional meaning of behavior is not inherent in behavior itself. Rather than assuming a one-to-one coherence between a specific behavior and its emotional meaning (such as a smile for happiness, or an eye-roll for contempt) it is assumed first, that the affective meaning of behavior is constructed in the context of the interaction itself, and second, that there are an infinite number of ways to express a certain affective quality in an interaction. A smile, for example, can be condescending, and an eye-roll can be seen as empathetic if a team member complains about a vendor and his colleague rolls his eyes in agreement. The condescending comment and the eye-roll form what Schegloff [43] called an adjacency pair as the interpretation of the colleague's eye-roll is directly contingent upon the teammate's complaint. In the same way, the emotional interpretation of backchanneling behavior is contingent upon its placement within the interaction. The slight nodding and looking by the listener when placed concurrently with the speaker as implemented in our study can transform backchanneling to convey positive listening and engagement.

Backchanneling from a social-functional perspective can be seen as an interactive device for regulating emotions in teams. Emotions are contagious. They easily spread within a team and it only takes one person's negative emotions to spoil an entire team [3, 21]. Since emotions are crucial determinants of team performance [4], this opens up the possibility that robots, through their backchanneling behavior, can help to regulate the emotions of a team and through that ultimately affect the team's performance. Understanding when and what robot affective displays are contagious will be important in designing more effective robot team members. At the extreme, robots' backchanneling and displays could be used to intentionally regulate the emotions of work teams, for example, diffusing stressful situations when the work is demanding and dangerous.

More broadly, it becomes important to consider the affective meanings and effects of the interactive capabilities currently being developed in the field of robotics (nodding, speech modulation, gaze, body orientation, etc). A social-functional perspective applied to these behaviors might reveal affective implications for behaviors that have been thought of as purely cognitive. For example, simple pointing gestures can have important positive functions in directing attention but they can also be interpreted as a domineering or even condescending gesture. Focusing on these social interpretations of behavior is particularly important since brief exposure to a nonverbal expression is enough for people to infer a complex image of a person's (and perhaps a robot's) personality [31] and capabilities.

In sum, our results confirm that subtle forms of backchanneling by robots may ease team functioning and improve perceptions of robots' engagement in complex tasks, but at the same time create more negative impressions of the robots' competence. These results suggest that more attention is needed to subtle affective cues and their role in human-robot teams.

ACKNOWLEDGMENTS

We thank our anonymous reviewers for their helpful comments on previous versions of this document. The authors gratefully acknowledge two grants by the Office of Naval Research that supported the work presented here (#N000140710749, #N000140910112).

REFERENCES

1. ARToolKit. *[Software]*. Available from: http://www.hitl.washington.edu/artoolkit/.
2. Aviezer, H., et al., *Angry, Disgusted, or Afraid?* Psychological Science, 2008. 19(7): p. 724-732.
3. Barsade, S.G., *The Ripple Effect: Emotional Contagion and its Influence on Group Behavior.* Administrative Science Quarterly, 2002. 47(4): p. 644-675.
4. Barsade, S.G. and D.E. Gibson, *Why does affect matter in organizations?* The Academy of Management Perspectives, 2007. 21(1): p. 36-59.
5. Bethel, C.L. and R.R. Murphy, *Affective expression in appearance constrained robots*, in *Proceedings of the 1st ACM SIGCHI/SIGART conference on Human-robot interaction*. 2006, ACM: Salt Lake City, Utah, USA. p. 327-328.
6. Bethel, C.L. and R.R. Murphy, *Non-facial/non-verbal methods of affective expression as applied to robot-assisted victim assessment*, in *Proceedings of the ACM/IEEE international conference on Human-robot interaction*. 2007, ACM: Arlington, Virginia, USA. p. 287-294.
7. Bickmore, T.W. and R.W. Picard, *Towards caring machines*, in *CHI '04 extended abstracts on Human factors in computing systems*. 2004, ACM: Vienna, Austria. p. 1489-1492.
8. Breazeal, C., et al. *Effects of nonverbal communication on efficiency and robustness in human-robot teamwork*. in *Intelligent Robots and Systems, 2005. (IROS 2005). 2005 IEEE/RSJ International Conference on*. 2005.
9. Breazeal, C. and B. Scassellati. *How to build robots that make friends and influence people*. in *Intelligent Robots and Systems, 1999. IROS '99. Proceedings. 1999 IEEE/RSJ International Conference on*. 1999.
10. Breazeal, C.L., *Designing sociable robots*. 2004: The MIT Press.
11. Bruce, A., I. Nourbakhsh, and R. Simmons. *The role of expressiveness and attention in human-robot interaction*. in *Robotics and Automation, 2002. Proceedings. ICRA '02. IEEE International Conference on*. 2002.
12. Brunken, R., J.L. Plass, and D. Leutner, *Direct Measurement of Cognitive Load in Multimedia Learning.* Educational Psychologist, 2003. 38(1): p. 53-61.
13. Campbell, D.J., *Task complexity: A review and analysis.* The Academy of Management Review, 1988. 13(1): p. 40-52.

14. CARMEN. *Robot Navigation Toolkit*. Available from: http://carmen.sourceforge.net/.
15. Casper, J. and R.R. Murphy, *Human-robot interactions during the robot-assisted urban search and rescue response at the World Trade Center.* Systems, Man, and Cybernetics, Part B: Cybernetics, IEEE Transactions on, 2003. 33(3): p. 367-385.
16. Cereproc. *[Software]*. Available from: http://www.cereproc.com/.
17. Clark, H.H. and S.A. Brennan, *Grounding in Communication*, in *Perspectives on socially shared cognition*, L.B. Resnick, J.M. Levine, and S.D. Teasley, Editors. 1991, APA Book.: Washington.
18. Coan, J.A. and J.M. Gottman, *The Specific Affect Coding System (SPAFF)*, in *Handbook of emotion elicitation and assessment.*, J.A. Coan and J.J.B. Allen, Editors. 2007, New York, NY, US: Oxford University Press. p. 267-285.
19. Dennis, A.R. and S.T. Kinney, *Testing Media Richness Theory in the New Media: The Effects of Cues, Feedback, and Task Equivocality.* Information Systems Research, 1998. 9(3): p. 256-274.
20. Earley, P.C., *Influence of information, choice and task complexity upon goal acceptance, performance, and personal goals.* Journal of Applied Psychology, 1985. 70(3): p. 481-491.
21. Felps, W., T.R. Mitchell, and E. Byington, *How, when, and why bad apples spoil the barrel: Negative group members and dysfunctional groups.* Research in organizational behavior, 2006. 27: p. 175-222.
22. Fincannon, T., et al. *Evidence of the need for social intelligence in rescue robots*. in *2004 IEEE/RSJ International Conference on Intelligent Robots and Systems*. 2004. Sendai, Japan: IEEE.
23. Fong, T., I. Nourbakhsh, and K. Dautenhahn, *A survey of socially interactive robots.* Robotics and Autonomous Systems, 2003. 42(3): p. 143-166.
24. Gockley, R., R. Simmons, and J. Forlizzi. *Modeling Affect in Socially Interactive Robots*. in *Robot and Human Interactive Communication, 2006. ROMAN 2006. The 15th IEEE International Symposium on*. 2006.
25. Jacob, M., et al., *Gestonurse: a robotic surgical nurse for handling surgical instruments in the operating room.* Journal of Robotic Surgery, 2012. 6(1): p. 53-63.
26. Johnson, C., *Gender, Legitimate Authority, and Leader-Subordinate Conversations.* American Sociological Review, 1994. 59(1): p. 122-135.
27. Jones, H. and P. Hinds. *Extreme work teams: using swat teams as a model for coordinating distributed robots*. in *ACM conference on Computer supported cooperative work, CSCW'02*. 2002. New Orleans, Louisiana, USA.: ACM.
28. Jung, M.F., J. Chong, and L.J. Leifer. *Group Hedonic Balance and Pair Programming Performance: Affective Interaction Dynamics as indicators of Performance*. in *ACM SIGCHI Conference on Human Factors in Computing Systems (CHI'12)*. 2012. Austin, Texas, USA.
29. Kahn, W.A., *Psychological Conditions of Personal Engagement and Disengagement at Work.* Academy of Management Journal, 1990. 33(4): p. 692-724.
30. Keltner, D. and J. Haidt, *Social functions of emotions*, in *Emotions: Currrent issues and future directions*, T.J.M.G.A. Bonanno, Editor. 2001, Guilford Press: New York, NY, US. p. 192-213.
31. Knutson, B., *Facial expressions of emotion influence interpersonal trait inferences.* Journal of Nonverbal Behavior, 1996. 20(3): p. 165-182.
32. Lee, K.M., et al., *Can Robots Manifest Personality?: An Empirical Test of Personality Recognition, Social Responses, and Social Presence in Human–Robot Interaction.* Journal of Communication, 2006. 56(4): p. 754-772.
33. Lewis, K., *Measuring transactive memory systems in the field: Scale development and validation.* Journal of Applied Psychology, 2003. 88(4): p. 587-604.
34. Liu, C., et al. *Generation of nodding, head tilting and eye gazing for human-robot dialogue interaction*. in *7th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*. 2012. Kyoto, Japan: IEEE.
35. Maynard, D.C. and M.D. Hakel, *Effects of objective and subjective task complexity on performance.* Human Performance;Human Performance, 1997. 10(4): p. 303-330.
36. Murphy, R.R., *Human-robot interaction in rescue robotics.* Systems, Man, and Cybernetics, Part C: Applications and Reviews, IEEE Transactions on, 2004. 34(2): p. 138-153.
37. Murphy, R.R. and J.L. Burke, *Up from the Rubble: Lessons Learned about HRI from Search and Rescue.* Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 2005. 49(3): p. 437-441.
38. Murphy, R.R., et al., *Emotion-based control of cooperating heterogeneous mobile robots.* Robotics and Automation, IEEE Transactions on, 2002. 18(5): p. 744-757.
39. Nass, C. and B. Reeves, *The media equation: How people treat computers, televisions, and new media as real people and places*. 1996, Cambridge University Press.
40. Parasuraman, R. and C.A. Miller, *Trust and etiquette in high-criticality automated systems.* Commun. ACM, 2004. 47(4): p. 51-55.

41. Pasupathi, M., et al., *Responsive Listening in Long-Married Couples: A Psycholinguistic Perspective.* Journal of Nonverbal Behavior, 1999. 23(2): p. 173-193.
42. Rose, R., M. Scheutz, and P. Schermerhorn. *Empirical investigations into the believability of robot affect*. in *Proceedings of the AAAI Spring Symposium.* 2008.
43. Schegloff, E.A., *Sequence organization in interaction: A primer in conversation analysis I*. Vol. 1. 2007: Cambridge Univ Pr.
44. Shah, J., et al., *Improved human-robot team performance using chaski, a human-inspired plan execution system*, in *Proceedings of the 6th international conference on Human-robot interaction*. 2011, ACM: Lausanne, Switzerland. p. 29-36.
45. Sidner, C.L., et al., *Explorations in engagement for humans and robots.* Artificial Intelligence, 2005. 166(1): p. 140-164.
46. Sidner, C.L., et al., *The effect of head-nod recognition in human-robot conversation*, in *Proceedings of the 1st ACM SIGCHI/SIGART conference on Human-robot interaction*. 2006, ACM: Salt Lake City, Utah, USA. p. 290-296.
47. Sphinx, C. *[Software]*. Available from: http://cmusphinx.sourceforge.net/.
48. Stubbs, K., P.J. Hinds, and D. Wettergreen, *Autonomy and Common Ground in Human-Robot Interaction: A Field Study.* Intelligent Systems, IEEE, 2007. 22(2): p. 42-50.
49. Takayama, L., V. Groom, and C. Nass, *I'm sorry, Dave: i'm afraid i won't do that: social aspects of human-agent conflict*, in *Proceedings of the 27th international conference on Human factors in computing systems*. 2009, ACM: Boston, MA, USA. p. 2099-2108.
50. van Gerven, P.W.M., et al., *Modality and variability as factors in training the elderly.* Applied Cognitive Psychology, 2006. 20(3): p. 311-320.
51. Vicon. *[Software]*. Available from: http://www.vicon.com/.
52. Wang, E., et al., *Effects of head movement on perceptions of humanoid robot behavior*, in *Proceedings of the 1st ACM SIGCHI/SIGART conference on Human-robot interaction*. 2006, ACM: Salt Lake City, Utah, USA. p. 180-185.
53. Wang, L. and C. Chen, *A Combined Optimization Method for Solving the Inverse Kinematics Problem of Mechanical Manipulators, .* IEEE Trans. On Robotics and Applications, 1991. 7(4): p. 489-499.
54. Watson, D., L.A. Clark, and A. Tellegen, *Development and validation of brief measures of positive and negative affect: The PANAS scales.* Journal of Personality and Social Psychology, 1988. 54(6): p. 1063-1070.
55. Weick, K.E., *The Vulnerable System: An Analysis of the Tenerife Air Disaster.* Journal of Management, 1990. 16(3): p. 571-593.
56. Yamazaki, A., et al., *Precision timing in human-robot interaction: coordination of head movement and utterance*, in *Proceedings of the twenty-sixth annual SIGCHI conference on Human factors in computing systems CHI'08*. 2008, ACM: Florence, Italy. p. 131-140.

APPENDIX: SURVEY ITEMS

Manipulation Checks

Back-channeling:
"I felt that the ground robots were looking at me" (7pt, "strongly disagree" – "strongly agree")

Task Complexity:
"I found this to be a complex task"
"The task required me to coordinate many different things at the same time". (7pt, "totally disagree" – "totally agree")

Team Functioning

Stress:
"Indicate to what extent you have felt distressed during the interaction task"
"Indicate to what extent you have felt upset during the interaction task"
"Indicate to what extent you have felt irritable during the interaction task" (5pt, "very slightly or not at all" – "extremely")

Cognitive Load:
"The task was mentally demanding" (7pt, "totally disagree" – "totally agree")

Perceived Team Coordination:
"Our team worked together in a well-coordinated fashion."
"Our team had very few misunderstandings about what to do."
"Our team needed to backtrack and start over a lot." (reverse scored)
"We accomplished the task smoothly and efficiently."
"There was much confusion about how we would accomplish the task." (reverse scored) (5pt, "strongly disagree" – "strongly agree")

Perceptions of robots' engagement

Perceived Distractedness
"I felt that the ground robots were distracted"
"I felt that the ground robots were unfocused" (7pt, "strongly disagree" – "strongly agree")

Perceived Responsiveness
"I felt that the ground robots were responsive" (7pt, "strongly disagree" – "strongly agree")

Perceptions of robots' competence

Perceived Intelligence
"How much did you feel as if you were accompanied by intelligent beings?" (5pt, "not at all" – "absolutely")

Perceived Competence
"The ground robots are very competent" (7pt, "strongly disagree" – "strongly agree")