Measuring Attitudes Towards Telepresence Robots

Katherine M. Tsui, Munjal Desai, Holly A. Yanco, Henriette Cramer, and Nicander Kemper

***Abstract*—Studies using Nomura et al.'s "Negative Attitude toward Robots Scale" (NARS) [1] as an attitudinal measure have featured robots that were perceived to be autonomous, independent agents. State of the art telepresence robots require an explicit human-in-the-loop to drive the robot around. In this paper, we investigate if NARS can be used with telepresence robots. To this end, we conducted three studies in which people watched videos of telepresence robots (n=70), operated telepresence robots (n=38), and interacted with telepresence robots (n=12). Overall, the results from our three studies indicated that NARS may be applied to telepresence robots, and culture, gender, and prior robot experience can be influential factors on the NARS score.**

***Index Terms*—Telepresence robots, human-robot interaction, Negative Attitude toward Robots Scale (NARS)**

1. INTRODUCTION

As robots become more commonplace in society, it is important to understand how people feel about them, including how they look, how they behave, and their purpose. For robots to be accepted by the masses, they must be designed in a fashion that would facilitate easier adoption. One of the initial steps is to find and to understand any preconceived notions or biases that people might have against robots so that they can then be addressed. Nomura et al. had a goal of creating a psychological scale to examine people's attitudes, anxiety, and assumptions about robots as a means to understand how people would react to robots in everyday situations [1]–[3]. After development and refinement, the Negative Attitudes toward Robots Scale (NARS) was created in 2003 [1].

Telepresence robots can be described as mobile embodied video conferencing systems with live two-way video and audio communication. In state of the art telepresence robots, there is an explicit human-in-the loop (i.e., the remote operator), who must manually drive the remote robot. To our knowledge, NARS has only been used with robots that are autonomous or appear to be autonomous through "Wizard of Oz" operation [4] (with the operator hidden from the participant) in short, scripted interactions (see Table II). Our goal was to evaluate if NARS could be extended to robots that were known to have a human-in-the-loop, specifically telepresence robots.

In this paper, we provide a survey of the use of NARS, focusing on how the scores have been used to explain human interaction behavior. We then discuss a series of studies in which we applied NARS to telepresence robots in an online video survey, an in-person study of people operating telepresence robots, and an in-person study of people physically present with a robot while interacting with a remote operator.

Manuscript received December 15, 2010.

K. Tsui, M. Desai, and H. Yanco are with the University of Massachusetts Lowell. H. Cramer is with SICS & Mobile Life Centre; part of her contribution to this paper was carried out during the tenure of an ERCIM "Alain Bensoussan" Fellowship Programme. N. Kemper is with KZA b.v.

2. NEGATIVE ATTITUDE TOWARD ROBOTS SCALE (NARS)

Nomura et al. hypothesized that people with high levels of communication apprehension towards people might also have communication apprehension towards robots since they posited that people do not discriminate between humans and agents with respect to communication. They developed a psychological tool to measure people's anxiety towards robots which evolved into NARS and its three subscales: "situations of interaction with robots" (NARS-S1), "social influence of robots" (NARS-S2), and "emotions in interaction with robots" (NARS-S3) [1], [6].

With NARS, participants are asked to rate the items shown in Table I on a scale from 1 (I strongly disagree) to 5 (I strongly agree) with 3 as "undecided." A higher score for NARS-S1 and NARS-S2 indicates a more negative attitude towards robots; conversely, a lower score indicates a more positive attitude. NARS-S3 is an inverse scale, so a higher score indicates a more positive attitude; conversely, a lower score indicates a more negative attitude.

2.1. *The Use of NARS*

NARS has been used in a range of scenarios since its creation in 2003, as shown in Table II. NARS has been used with populations from several countries and with different robots. Studies have shown that several factors affect the NARS subscale scores, including gender [5], [7]–[9]; age [5]; prior robot experience [8]; and culture [8], [10]. (A detailed survey of these studies is provided in [11].)

The goal of NARS is to predict user interaction based on people's negative attitudes towards robots. Preliminary evidence of the usefulness of NARS can be found in [5], [7], [9], [12]. Nomura et al. conducted a series of experiments in which the participants were asked to interact with Robovie. In [5], the interaction involved simply talking to the robot. The authors found that participants with higher NARS-S1 scores took longer to initially talk to the robot than those participants that had lower S1 scores. This finding indicated that participants with more negative attitudes towards robots on the interaction subscale took longer to initiate their interaction with a robot.

Increasing the interaction complexity, participants in [7] were first asked to talk with Robovie and, at the end, the robot asked the participants to touch it. Prior to the interaction, the participants were asked to answer NARS and were also asked if they had prior experience communicating with robots. The authors noticed differences in the behavior of participants based on their NARS scores. They divided the participants into two groups based on their NARS score. The elapsed time from when participants entered the room and talked to the robots was less for participants that had lower NARS-S1 (interaction)

TABLE I
NEGATIVE ATTITUDE TOWARD ROBOTS SCALE (NARS)

Subscale	Items
NARS-S1: Interaction	• I would feel uneasy if I was given a job where I had to use robots. • The word "robot" means nothing to me. • I would feel nervous operating a robot in front of other people. • I would hate the idea that robots or artificial intelligences were making judgements about things. • I would feel very nervous just standing in front of a robot. • I would feel paranoid talking with a robot.
NARS-S2: Social	• I would feel uneasy if robots really had emotions. • Something bad might happen if robots developed into living beings. • I feel that if I depend on robots too much, something bad might happen. • I am concerned that robot would be a bad influence on children. • I feel that in the future society will be dominated by robots.
NARS-S3: Emotion (inverse)	• I would feel relaxed talking with robots. • If robots had emotions, I would be able to make friends with them. • I feel comforted being with robots that have emotions.

Items from Nomura et al. [1] translated by Bartneck and colleagues [5].

scores than those that had higher score. This finding indicated that the tendency to avoid interacting with robots could be predicted based on NARS.

A similar study was conducted in [9]; participants entered a room that had only the robot (Robovie), were prompted to verbally interact with the robot after 30 seconds, and finally were prompted to touch the robot after 30 seconds. Prior to the interaction, the participants completed NARS. The time elapsed from the participants being initially encouraged to touch the robot and actually touching the robot was negatively influenced by the NARS-S2 (social) subscale. When looking for a gender difference, the authors found a positive influence between the time that male participants took to talk to the robot after entering the room and the NARS-S1 (interaction) subscale; the time that male participants took to touch the robot was positively influenced by the NARS-S3 (emotion) subscales and negatively influenced by the NARS-S2 (social) subscale. The study also found some differences for the female participants. They found that the distance at which that the female participants stood from the robot was positively influenced by the NARS-S1 (interaction) subscale and negatively by the NARS-S2 (social) subscale.

Finally, Nomura et al. [12] conducted an experiment to investigate the relationship between anxiety and negative attitudes towards robots and the allowable interaction distance to robots. Robovie-M had two walking speeds (slow at 6 cm/s or fast at 12 cm/s) and would move towards a participant until the participant asked the experimenter to stop the robot. The authors found a trend between the allowable distance and the NARS-S2 (social) subscale.

Recently, NARS has started to gain wider use by HRI researchers as a supplemental attitudinal measure to help explain some of the differences that they observe by partitioning participants into groups based in their NARS score, as in [7]. This use of NARS is described in a 2009 study by Cramer et al. [14], [15]. They conducted an experiment in which they examined how people perceived a robot when physical contact and help style were factors. In the between-subjects online survey, participants were asked to watch one of four videos featuring a Robosapien helping a woman using a word processing application on a computer. When the woman has a computer problem, the robot gives advice as to what she should do to recover her work. In the condition with physical contact, the woman taps the Robosapien to get its attention for help, while the Robosapien touches the woman on the shoulder. The video also shows the woman and the Robosapien sharing a hug and high-fiving each other at the end of the video. In the condition without the physical contact, the woman and the Robosapien only converse. The style of the Robosapien helping the woman was also manipulated; the robot either helps when asked (reactive) or offers advice (proactive).

Participants were asked a subset of the NARS-S1 subscale questions with the original wording from [12] and modified S3 subscale questions in which two of the questions were altered from their original wording. Participants were divided into those having a positive attitude towards robots (overall NARS score below $\bar{x}$=3.4, SD=1.0) and those with negative attitudes (NARS score above $\bar{x}$). Participants with a more negative attitude towards robots perceived the robot in the video as more machine-like and less human-like. They also thought the robot had less empathic abilities, was less dependable and less credible. They assessed the human-robot relationship as less close. Attitudes towards robots did not interact with effects of empathic accuracy or situational valence.

Also in 2009, Syrdal et al. [16] conducted an experiment with 28 participants from the University of Hertfordshire. The participants included both students and staff, with 14 female and 14 male participants. The robot, a PeopleBot from MobileRobots, had two different behaviors: socially ignorant and socially interactive. For the task, participants were required to physically interact with the co-located robot. The room was made to resemble a living room. The tasks included moving around the room and interacting with the robot to get a pen. The robot had 2 sets of pre-defined behaviors with respect to the path that it took, the way it moved around the participant, and its speed. The participants were asked to evaluate the behavior of the robots. The authors found that NARS was useful in explaining the differences in behavior of the participants.

In 2010, Riek et al. [17] recruited 16 participants from a university in the United Kingdom. Of the 16 participants, 9

TABLE II
CHRONOLOGY OF STUDIES USING NARS

Study	Year	n	Language	Robot	Description
Nomura et al. [1]	2003	39 263 240	Japanese	-	Initial efforts to create a psychological scale for NARS
Nomura et al. [5]	2004	240	Japanese	Robovie	Investigated gender and prior experience differences from NARS
Bartneck et al. [10]	2005	96	English	-	Investigated effects of cultural differences from NARS
Nomura et al. [7]	2006	53	Japanese	Robovie	Investigated the relationship between peoples' negative attitudes and their interaction with robots
Nomura et al. [13]	2006	400	Japanese	-	Attempted to find the relationship between the participants assumptions and attitudes towards robots
Bartneck et al. [8]	2007	467	English	Aibo	Investigated effects of cultural differences and prior robot experience
Nomura et al. [12]	2007	17	Japanese	Robovie-M	Investigated the relationship between peoples' attitudes and anxiety and robot's proximity to the person towards and the distance maintained during interaction
Nomura et al. [9]	2008	38	Japanese	Robovie	Investigated elapsed time for verbal and physical interactions relating to peoples' attitudes and anxiety and their behavior towards robots
Cramer et al. [14], [15]	2009	119	English	Robosapien	Investigated the relationship between the robot's behavior involving contact with a person and people's attitudes
Syrdal et al. [16]	2009	28	English	Peoplebot	Used NARS to explain the participants' behaviors
Riek et al. [17]	2010	16	English	BERTI	Examined videos of humanoid gestures (beckon, give, and handshake)

were female and 7 male. They conducted a within-subjects study to determine how people react to gestures made by BERTI, a humanoid torso robot. They used NARS in their experiment since NARS had been verified with British participants, particularly from universities. (See Riek et al. [17] for full details about the participants' highly varied backgrounds.) The authors showed the participants 12 different videos where the gesture types, style, and orientations were different. With respect to NARS, participants with negative attitudes towards robots were found to be less adept at understanding gestures.

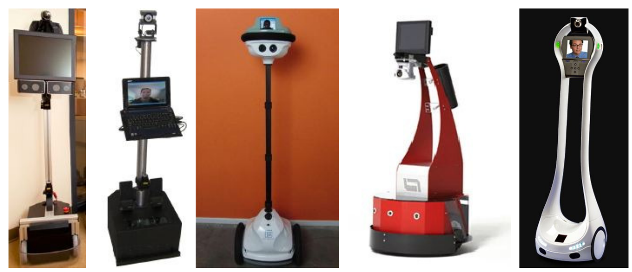

Fig. 1. Telepresence robots shown in videos (left to right): Willow Garage's Texai, SuperDroid's RP2W, Anybots' QB, RoboDynamics' Tilr, and VGo Communications' VGo. (Not to scale)

3. Case Study: Telepresence Robots

In all of the studies in Table II, the participants perceived the robots to be autonomous, independent agents. In this paper, we examine if NARS can also be used for telepresence robots which clearly have a human involved in the operation of the robot. In state of the art telepresence robots, an operator must log into a robot and manually drive the robot around.

To this end, we conducted an online video survey of five telepresence robots (Section 3.1), an in-person study of people operating two telepresence robots (Section 3.2), and an in-person study of people interacting with an experimenter who was operating a telepresence robot (Section 3.3). For each study, we calculate the consistency of NARS using Cohen's kappa (see the Appendix for details about Cohen's kappa calculation). We also investigate if there were any factors (i.e., gender, age, prior robot experience) that affected the NARS score, and examine if the score affect participant's impressions of and experiences with telepresence robots.

3.1. Study 1: Online survey

Using Amazon's Mechanical Turk [18], we conducted an online video survey of five telepresence robots in a between-subjects experiment with 80 participants. In the survey, participants were asked to provide their demographic information, including age, gender, occupation, country of citizenship, and pet ownership. We also solicited information about their prior robot experience and their video game usage in the last 12 months. They then completed baseline NARS rating. Participants were asked to rate the NARS statements on a scale from 1 (I strongly disagree) to 5 (I strongly agree) with 3 as "undecided."

Participants were asked to view three videos. After each video, the participants scored the NARS questions; they also provided their impressions of the robot's anthropomorphism [19], [20], likeability [19], [21], familiarity [19], and eeriness [19] (see Table III). In the first video, participants saw one of four of the Robosapien videos (chosen at random) used in Cramer et al.'s study [14], [15]. The second video was of the Anybots' QB robot.[1] The remaining video showed one of four telepresence robots: Willow Garage's Texai,[2] VGo Com-

[1] http://tinyurl.com/youtube-anybots-qb
[2] http://tinyurl.com/youtube-willow-garage-texai

TABLE III
IMPRESSIONS OF THE TELEPRESENCE ROBOT (SEMANTIC SCALE FROM 1 LEFT-MOST TO 7 RIGHT-MOST)

Likeability [19], [21]	awful/nice unfriendly/friendly unkind/kind unpleasant/pleasant
Anthropomorphism [19], [20]	fake/natural machine-like/human-like unconscious/conscious artificial/lifelike moves rigidly/moves elegantly
Familiarity [19]	strange/familiar
Eeriness [19]	extremely eerie/not at all eerie

TABLE IV
STUDY 1: CRONBACH ALPHA VALUES OF NARS SUBSCALES (n_{female} = 29, n_{male} = 41) (GRAY INDICATES "POOR" OR "UNACCEPTABLE.")

		S1	S2	S3
Baseline	Overall	0.61	0.63	0.80
	Female	0.58	0.64	0.82
	Male	0.63	0.61	0.79
After Robosapien video	Overall	0.69	0.70	0.81
	Female	0.64	0.69	0.87
	Male	0.71	0.72	0.73
After telepresence robot videos	Overall	0.77	0.78	0.80
	Female	0.71	0.83	0.82
	Male	0.80	0.75	0.78

munications' VGo,[3] RoboDynamics' TiLR,[4] or SuperDroid's RP2W.[5] We created four versions of the survey for each of the remaining telepresence robots; twenty participants saw each video. We randomized whether the participants saw the QB robot as the second or third video. All videos of the telepresence robots demonstrated communication capabilities through the robot. All videos except RP2W showed people interacting with the operator through the robot; if the operator's live video was not shown on the robot's screen, the operator's interface was shown.

The average time spent on this survey was 1 hour 59 minutes (SD=3 hours 24 minutes). The median time was 52 minutes. It should be noted that due to the length of the survey, we allowed participants to take up to 24 hours to complete the survey but asked that they complete the survey in a single session. Participants were compensated \$1.50 for completing the survey.

3.1.1) Participants: Seventy participants provided usable data. We discarded 10 participants' data due to submission of duplicate surveys, exiting the survey before completion, or incorrectly answering validation questions. Participants' ages ranged from 18 to 59 years ($\bar{x}$=28.7 years, SD=8.7); forty-one participants were male and twenty-nine female. Twenty-nine participants reported the United States as their country of citizenship. Thirty-one participants were citizens of India, and the remaining were citizens of Australia (1 participant), Canada (2), Greece (1), Honduras (1), Russia (1), Singapore (1), Switzerland (1), and United Kingdom (1). There were no instances of dual citizenship.

3.1.2) Consistency: For all seventy responses, we computed the Cronbach alpha values of the NARS responses after each of the videos (shown in Table IV). First, participants were asked for their ratings in the demographic information section. They were asked for their ratings again in the replication of the Robosapien videos. Lastly, we aggregated the scores for all of the telepresence robots. We found that the participants had the least consistent scores for the NARS-S1 and S2 subscales (interaction and behavior, respectively) in the baseline section before a robot was presented to them; this finding is not unexpected because the participants could provide scores based only on their experience or inexperience with real and fictional

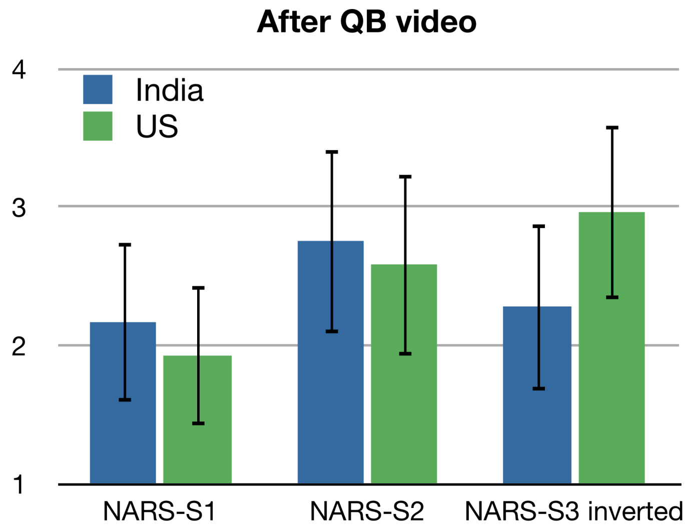

Fig. 2. Study 1: NARS ratings after watching video of QB robot (n_{India}=31, n_{US}=29). NARS-S3 has been corrected so that a lower value indicates a more positive score. P-values are provided for significant differences found.

robots. However, after viewing the Robosapien video, the participants' scores for the NARS-S1 and S2 subscales moved from a borderline acceptable alpha value (α>0.69) and then to a solidly acceptable value after viewing the telepresence robot videos (α>0.77). For the NARS-S3 subscale (emotion), the Cronbach alpha values would be classified as good to excellent reliability according to George and Mallory [22].

3.1.3) Results: Our analysis focuses on the QB robot as all seventy participants viewed the QB video, and the QB robot was to be used in Study 2. Sixty-eight of the seventy participants provided NARS ratings after watching the Anybots' QB demonstration video. We calculated the NARS scores by first inverting the NARS-S3 (emotion) subscale to correspond to the S1 and S2 subscales; a low S3 value now meant that the participant had a positive attitude towards robots. Gender and prior robot experience did not affect any of the NARS subscales. We investigated cultural differences affecting the NARS scores due to the two larger groups from India (n=31) and the US (n=29); however, two participants from the US did not provide NARS ratings after watching the QB video. Averaging NARS-S3 (emotion) subscale scores together, we found that the Indian participants had more positive views about robots ($\bar{x}$=2.28, SD=1.21) than the participants from the US ($\bar{x}$=2.96, SD=1.26) (p<0.01, t(59)=2.54 using a two-

[3] http://tinyurl.com/boston-globe-vgo
[4] http://tinyurl.com/youtube-robodynamics-tilr
[5] http://tinyurl.com/youtube-superdroid-rp2w

tailed unpaired t-test with unequal variance). There was no significant difference between the NARS-S1 and S2 subscale ratings for the US and Indian participants.

Thirty-six participants had an overall positive NARS score (less then or equal to $\bar{x}_{NARS\,overall}$=2.43, SD=1.29). We found that participants with positive NARS scores rated the QB higher ($\bar{x}$=5.26, SD=1.52) than the participants with negative NARS scores ($\bar{x}$=4.43, SD=1.36) with respect to its eeriness; the difference was weakly significant (p<0.06, using a Wilcoxon-Mann-Whitney test). Aside from eerieness, we did not find significant differences between the the participants' reported impressions (from Table III).

3.2. Study 2: Controlling telepresence robots

In Study 1, we used NARS with videos of teleoperated robots. Our next step was to apply NARS with people who were operating telepresence robots. We conducted a between-subjects, in-person study in which participants were asked for their initial impressions of a telepresence robot. Participants were provided with an overview of the study and an informed consent form. After providing their consent, participants completed a demographic survey including age, gender, occupation, and computer usage and expertise. They then completed baseline NARS rating. We also solicited information about their prior robot experience and their video game usage in the last 12 months.

Participants then operated one of two telepresence robots: an Anybots' QB [23] or a VGo Communication's VGo [24]; descriptions of the robots are provided in Table V. (It should be noted that the QB and the VGo robots were in beta and alpha testing respectively. Both began selling in Fall 2010.) The robot and associated interface was described by a test administrator, and the participants learned how to remotely operate the robot in a hands-on training session. Training ended when the participant felt comfortable operating the robot.

There were three tasks in this experiment. The first was a navigation task through an office environment. The robot was placed at a starting location inside of a group cubicle, away from the participant. The participant was instructed to meet the second test administrator in a specified conference room. The second task was a communication task. Once inside the conference room, the participant engaged in a conversation with the test administrator who was sitting at a table. The third task involved viewing a dry erase board. A map of a sub-section of campus was drawn on a 4 foot high by 8 foot wide dry erase board. The participant was asked to provide instruction on how to drive the robot between two buildings (not connected or adjacent to each other).

After completing these tasks, the participant completed a post-experiment survey which included the NARS rating and items to gauge the participant's impressions on the robot's anthropomorphism, likeability, familiarity, and eeriness (Table III). The post-survey experiment also contained 7-point semantic differential scale statements relating to the overall system and perceived appearance of the experimenter. The participant was debriefed by the test administrator and was shown the robot in person. The average length of a session from the overview of the experiment through debriefing was one hour.

3.2.1) Participants: This experiment was conducted at Google in Mountain View, CA during July and August 2010. Forty-one people who were full-time employees of Google participated in this experiment. Thirty-three participants successfully operated the robot. The remaining participants experienced technical difficulties and therefore did not complete the post-experiment survey; for the purposes of this analysis, we excluded their pre-experiment data.

Of the thirty-three participants who successfully drove the robot, twenty-three were male and ten were female. Eighteen participants reported their occupation as an engineer. Occupations of the fifteen remaining participants included program manager, financial analyst, product marketing manager, researcher, system administrator, researcher, and customer support. The average age of the participants was 30.6 years (SD=6.5) with a range of 23 to 49.

All participants had extensive experience with computers. Participants reported using a computer at work on average 44.3 hours per week (SD=7.9) and 19.7 hours per week (SD=13.0) in their free time. Seven participants reported their computer expertise as moderate, ten as experts, and thirteen as gurus. Participants reported using several different computer platforms: Macintosh (24 participants), PC (17), and Unix or Linux (21). Thirteen of thirty-three participants reported prior experience with robots. Twenty participants reported that they played video games; these people played an average of 3.3 hours per week. Thirteen of them played real time strategy games, and nine of the twenty played first person shooter video games.

3.2.2) Consistency: For the thirty-three participants, we computed the Cronbach alpha values for their baseline NARS response from the pre-experiment survey and also for their responses provided in the post-experiment survey after having driven a telepresence robot (shown in Table VI). Fourteen participants used the QB robot (Group 1, or G1), and nineteen used the VGo (Group 2, or G2). As shown in Table VI, we noticed that the baseline NARS-S3 subscale (emotion) borders on the edge of the "questionable" category with respect to consistency (α=0.58). We further looked at G1 versus G2 and found that G1 has an acceptable level of consistency with α=0.70, and G2's α=0.48 falls into the "unacceptable" category. The inconsistency in G2's baseline NARS-S3 ratings revealed a more interesting discrepancy.

Participants overall rated the statement "I would feel relaxed talking with robots" (NARS.S3.1.inverted) significantly lower ($\bar{x}$=2.39, SD=0.97) than the other two statements in the baseline NARS-S3 subscale. The statements "if robots had emotions, I would be able to make friends with them" (NARS.S3.2.inverted) averaged 3.00 (SD=0.94), and "I feel comforted being with robots that have emotions" (NARS.S3.3.inverted) averaged 3.42 (SD= 0.83). Using two-tailed paired t-tests, we found this difference to be signif-

[6]At the time of this study, the QB robot driver could view live video from the robots cameras, but the screen on the head of the QB robot was blank. Since this study, the QB robot shows a profile picture of the robot driver. Two-way video is planned.

TABLE V
KEY FEATURES OF THE ANYBOTS' QB (LEFT) AND THE VGO ROBOTS (RIGHT) USED IN STUDY 2

	QB [23]	**VGo** [24]
Unit cost	$15K	$5K
Drive	2 wheels (dynamically balancing)	2 wheels and 2 casters
Top speed	3.5 mph	2.75 mph
Height	3'2" to 6'3" (manually adjusted)	4'
Weight	35 lbs	18 lbs
Battery life	4-6 hours	6 or 12 hour battery option
Microphones	3 on top of head (equally spaced)	4 around video screen (2 front, 2 back)
Speakers	1 on top of head	2 (woofer in base, tweeter in head)
Volume control (robot side)	no	yes, when in a call
Screen size	3.5" diagonal	6" diagonal
Number of cameras	1 front facing and 1 facing down	1 front facing
Camera tilt	no (fixed)	180 degrees
Deictic reference	yes (laser pointer)	no
Occupancy indicator	blue LEDs around eyes	red and green LEDs on sides of screen
Operating systems	MacOS with Firefox 3.6	Windows 7/Vista/XP
Navigation control	keyboard (arrow keys or WASD)	mouse "Click and Go" or arrows keys
2-way audio	yes	yes
2-way video	no (planned feature)[6]	yes
WiFi access point switching	no (planned feature)	yes

TABLE VI
STUDY 2: CRONBACH ALPHA VALUES OF NARS SUBSCALES (GRAY INDICATES "POOR" OR "UNACCEPTABLE.")

	Baseline			*After telepresence robot*		
	Overall	G1	G2	Overall	G1 (QB)	G2 (VGo)
S1	0.72	0.79	0.67	0.65	0.80	0.64
S2	0.62	0.66	0.63	0.74	0.62	0.76
S3	0.58	0.70	0.48	0.66	0.67	0.65

icant (p<0.01 with t(32)=2.63 and 5.83 respectively). We hypothesize that this higher average rating for the statement "I would feel relaxed talking with robots" (NARS.S3.1.inverted) may be enhanced by the participants' extreme familiarity with technology including robots.

Aside from the NARS-S3 subscale being more inconsistent than expected, the remaining NARS subscale ratings from the pre-experiment survey appear to be consistent. When using NARS as an attitudinal measure, respondents may draw upon any previous robot experience (ranging from fictional robots in cartoons and science-fiction movies to service robots used in manufacturing, car washing, or vacuuming), which may be the case particularly when used as a pre-experiment or pre-interaction item. The results after the use of the telepresence robots were consistent. It is likely the participants were likely envisioning the robot they had just used when they filled out the post-experiment NARS ratings. In Study 3, we only administered NARS after the interaction with the robot.

3.2.3) Results: Using two-tailed unpaired t-tests with unequal variance on the NARS ratings from the baseline and then after using the telepresence robot, we did not find any significant differences in the population overall between the two types of robots.

Like Bartneck et al. [8], we also found a gender difference. Bartneck et al. found that females were more positive than males in the NARS-S2 subscale (social); however, gender specific data was not directly reported [8]. In our study, we found that that females provided higher ratings (and therefore were more negative) than males for all of the NARS subscales for both the baseline and post-experiment survey (Fig. 3). The female participants had more negative ratings for all of the NARS subscales with the exception NARS-S3 (emotion) for the baseline. After using the telepresence robots, the female participants had higher ratings for NARS-S1 (interaction) and NARS-S2 (social) subscales than the male participants. It should be noted that there was a small sample size for females (n=10) in this study with five using the QB robot and five using the VGo robot.

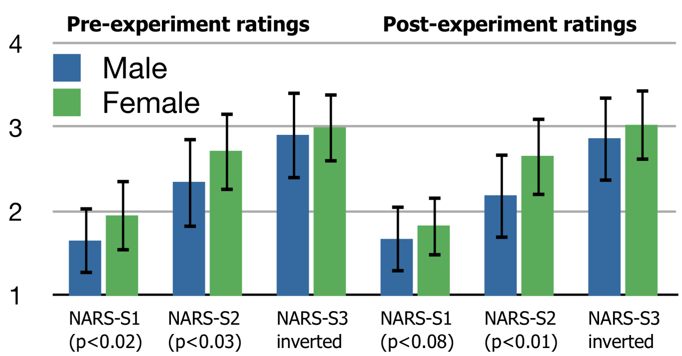

Fig. 3. Study 2: NARS ratings by gender (n_{female}=10, n_{male}=23). P-values are provided for significant differences found.

Seventeen participants had an overall positive NARS score which was less then or equal to the average ($\bar{x}_{NARS\,overall}$=2.20, SD=1.03). Using $\bar{x}_{NARS\,overall}$ to partition the participants, we found that participants with positive attitudes towards robots rated the telepresence robots significantly higher ($\bar{x}$=5.72, SD=1.64, median=6) than the participants with negative attitudes ($\bar{x}$=4.60, SD=1.24, median=4) with respect to its eeriness (p<0.04, using a Wilcoxon-Mann-Whitney test). This finding is consistent with Study 1. Further, we find that the participants' overall NARS score correlated to their reported eeriness rating (r=-0.38); that is, a lower overall NARS score indicated a more positive attitude and rated the robot as less eerie. Aside from eeriness, no significant differences were found for the remaining impression items from Table III or the other semantic differential scale questions asked in the post-experiment survey.

3.3. *Study 3: Interacting with telepresence robots*

We continued work in investigating the appropriateness of NARS from the perspective of the people interacting with the operator through the robot. We conducted an in-person pilot study in which twelve participants from the University of Massachusetts Lowell interacted with another person who used a VGo telepresence robot. We designed a session in which the participate would engage with a person through a telepresence robot, emphasizing both the conversational ability and the mobility of the robot itself.

First, the participant was provided with an overview of the study, an informed consent form, and an optional video consent form by the first experimenter. After the participant provided his or her consent, the first experimenter called another person (the second experimenter[7]) using the VGo robot. The second experimenter drove the robot off its charger, turned toward the participant, and introduced herself. She offered water and snacks to the participant by turning and driving toward the snacks. She then drove over to the conference table at which the participant and the first experimenter were sitting.

The first experimenter then described a desert survival task modified from [25]–[27]:

> It is approximately 10am in mid-July and you have just crash landed in the Sonora Desert, Southwest USA. Your light twin-engine plane, containing the bodies of the pilot and the co-pilot, has completely burnt out, only the frame remaining. None of the rest of you have been injured.
>
> The pilot was unable to notify anyone of your position before you crashed. However, ground sightings taken shortly before the crash suggested that you are about 65 miles off-course from your originally filed flight plan. A few moments before the crash, the pilot indicated that the nearest known habitation was a mining camp 70 miles away in a south south-west direction (2 day walk). The immediate area is has minor elevation change and has occasional cactus, desert animals (such as coyotes, vultures, snakes, lizards, jack rabbits, and big horned sheep), and tumbleweed. The last weather report indicated that the temperature would reach 110 degrees F during the day, which means that the temperature within a foot of the surface will be 130 degrees F. The temperature at night would be in the single digits.
>
> You are dressed in lightweight clothing: short sleeved shirts, shorts, socks and leather shoes. Everyone has a handkerchief. Collectively your pockets contain \$1.53 in change, \$43 in notes, and 1 liter of water each.

The participant was given a list of four pairs of items and asked to choose what he or she would have wished to pack in his or her travel bags the day before. The first experimenter informed the participant that he or she should make the initial selections individually and then would have 15 minutes to discuss his or her choices with the second experimenter to come to a final consensus.

The first experimenter remained in the room while the participant made his or her initial selections. When the initial section was complete, the first experimenter gave the participant an envelope containing a copy of the scenario, a sheet to mark the final selections, and pictures with descriptions of each of the objects. Then the first experimenter left the room, noting his return in 15 minutes.

The participant and the second experimenter interactively discussed the selections for the above desert survival task. The pairs of objects were 1) two bananas vs. one packet of peanuts, 2) an emergency car blanket vs. a red and white parachute, 3) a knife vs. pistol, and 4) a map and compass vs. matches and a book. The order in which the objects were selected and discussed was randomized before the experiment began. The second experimenter followed a script in which she disagreed with the participant in order to create interactive discussion as in [26]. To encourage a two-way conversation, she disagreed with the participant about the choice of the bananas or the peanuts and also about the choice of the compass and map or the matches and book. Table VIII in the Appendix shows the pros and cons for each item.

After a consensus had been reached, the second experimenter turned and drove toward a printer on a nearby desk. She asked the participant to take the five page post-experiment survey and begin to complete it while she returned the robot to the charger. She thanked the participant for his or her time.

The first participant returned to the room while the participant completed the post-experiment survey, which included the NARS questions and demographic information of age, gender, prior robot experience, and prior video conferencing experience. The first participant then answered any questions the participant had relating to the experiment and presented him or her with a \$10 gift card. The average duration of a session was 45 minutes.

3.3.1) Participants: Twelve people participated in this study (4 female, 8 male). Eleven were students. Participants' ages ranged from 18 to 53 ($\bar{x}$=26.1 years, SD=10.3); one participant did not report age. Eight of the twelve participants reported prior experience with robots. Three participants reported having knowledge of appropriate behavior for the scenario based on family residing in the desert, watching "Survivor Man," or having been in the US Army for 3 years.

3.3.2) Consistency: For the twelve participants, we computed the Cronbach alpha values for their post-experiment

[7]The participant was not explicitly told that the person with whom he or she would interact was also an experimenter. This knowledge may have influenced the participant's item selection reasoning if he or she thought that the experimenter had existing knowledge about the desert survival scenario.

TABLE VII
STUDY 3: CRONBACH ALPHA VALUES OF NARS SUBSCALES AFTER INTERACTING WITH THE TELEPRESENCE ROBOT (GRAY INDICATES "POOR" OR "UNACCEPTABLE.")

	Overall	Prior robot experience (n=8)	No prior experience (n=4)
S1	0.77	0.85	0.43
S2	0.74	0.69	0.58
S3	0.59	0.66	0.25

NARS responses. We note that there may be an effect of prior robot experience affecting participants' reported NARS score as the participants who had prior robot experience had more consistent ratings than participants with no prior robot experience (shown in Table VII).

We note that the degree of direct interaction may have influenced the overall consistency of the NARS ratings. In Study 2, the participant operated the telepresence robot, which was a direct interaction. In Study 3, the participant was in the room with the telepresence robot and discussed the selection of objects with the telepresence operator, which is a less direct interaction than operating the robot. One participant in this study noted that she thought "the robot was going to do something." Thus, participants with no prior robot experience may have considered their knowledge of real and fictional robots in addition to or instead of the telepresence robot with which they had just interacted.

Also as in Study 2, we noticed that the NARS-S3 (emotion) subscale borders on the edge of the "questionable" category again with respect to consistency (α=0.58). Participants overall rated the statement "I would feel relaxed talking with robots" (NARS.S3.1.inverse) lower ($\bar{x}$=2.00 overall, for participants with prior robot experience, and participants with no prior experience, $SD_{overall}$=0.85, $SD_{experience}$=1.07, $SD_{noExperience}$=0) than the other two statements in NARS-S3. The statements "if robots had emotions, I would be able to make friends with them" (NARS.S3.2.inverse) averaged 2.50 (SD=1.09), and "I feel comforted being with robots that have emotions" (NARS.S3.3.inverse) averaged 3.17 (SD= 0.94). We hypothesize that this higher average rating for the statement "I would feel relaxed talking with robots" (NARS.S3.1.inverse) may be enhanced by the participants' familiarity with technology. NARS-S1 (interaction) and NARS-S2 (social) showed good overall consistency ($\alpha > 0.7$).

3.3.3) Results: We note that there may be an effect of prior robot experience affecting participants' reported NARS score in addition to the consistency rating. Fig. 4 shows that participants with prior robot experience gave different ratings than participants with no prior experience. Even with such a small sample size, we find that the participants with prior robot experience have lower NARS-S2 average scores ($\bar{x}$=2.5, SD=1.08) and therefore more positive attitudes than participants with no prior experience ($\bar{x}$=3.05, SD=1.05) (p=0.06, $t(10)$=2.09 using a two-tailed unpaired t-test with unequal variance). Four people had an overall positive NARS score (less than or equal to $\bar{x}_{NARS\,overall}$=2.12, SD=1.03), but we did not find any significant differences between the participants' reported impressions (from Table III) using $\bar{x}_{NARS\,overall}$ to partition

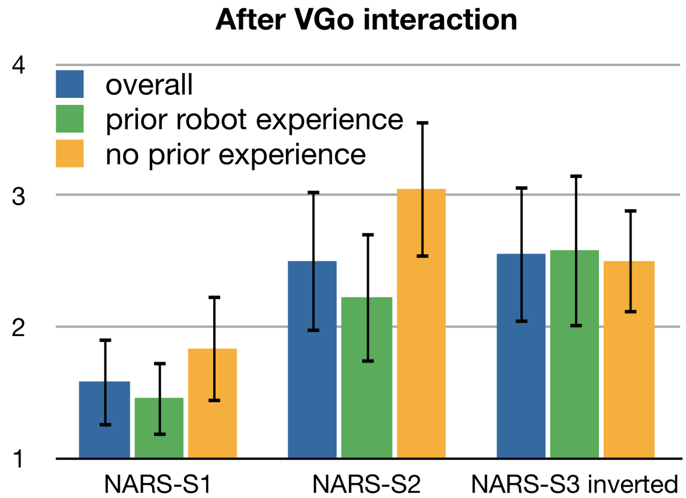

Fig. 4. Study 3: NARS ratings (overall: n=12, prior robot experience: n=8, no prior experience: n=4)

the participants into two groups.

4. DISCUSSION AND FUTURE WORK

In using NARS with telepresence robots, we have also found that culture (Study 1), gender (Study 2), and prior robot experience (Study 3) can influence the NARS scores, which is consistent with existing research [5], [7]–[10].

In Studies 2 and 3, we have exposed a potential need for updating the NARS-S3 subscale (emotion). In both studies, the NARS.S3.1 statement ("I would feel relaxed talking with robots") was rated significantly higher and the other two subscale items. For Study 2, we posited that the inconsistency may have been caused by participants drawing from their real and fictional knowledge of robots. The inconsistency remained after changing the experimental protocol to administering NARS only after the interaction with the robot. While the concept of a telepresence robot is not new [28], there has been an increase in telepresence robot platforms which are commercially viable over the last few years (e.g., RoboDynamics' Tilr, Anybots' QB, VGo Communication's VGo). These mobile video-chat robots do not yet have the popularity and thus presence of home vacuuming robots. If the function and need for these telepresence robots can be imagined by a layperson, particularly given their recent media coverage (i.e., New York Times [29], CNN [30], NBC [31]), then it may not be difficult for people to imagine talking with robots in general. Thus, the NARS.S3.1 statement ("I would feel relaxed talking with robots") may need to be replaced or reworded to keep the consistency of the NARS-S3 subscale at an acceptable level.

We used the $\bar{x}_{NARS\,overall}$ value to partition the participants in Studies 1 and 2 into two groups: those with positive attitudes towards robots having a NARS score less than or equal to $\bar{x}_{NARS\,overall}$, and the remaining as having negative attitudes. In Study 1, we found that participants with positive NARS scores rated the QB higher ($\bar{x}$=5.26, SD=1.52, median=5) than the participants with negative NARS scores ($\bar{x}$=4.43, SD=1.36, median=4.5) with respect to its eeriness. In Study 2, we found that participants with positive attitudes towards robots also rated the the telepresence robots significantly higher ($\bar{x}$=5.72, SD=1.64, median=6) than the participants with negative attitudes ($\bar{x}$=4.60, SD=1.24, median=4). It should be noted that in Study 2, there was an even distribution of the type of telepresence robot. For the seventeen people who had positive attitudes, 10 used a VGo robot and 7 used a QB robot; for the remaining twenty-one people who had negative attitudes, 14 used a VGo and 7 used a QB. This use of NARS provides insights into the consistency of the eeriness ratings as they are comparable between Study 1 and Study 2. It should be noted that both sets of participants were quite technical given the inherent recruiting requirements. Study 1 participants were existing users of MTurk and therefore had access to a computer and the Internet; Study 2 participants were full-time employees of Google.

While the performance metrics community traditionally has focused on quantitative data, we should take care not to try to capture user attitudes and experiences in rigid, numerical criteria only. Complementing scale-based methods with ethnographic techniques, more open observation, and interviews, for example, can yield insights as to why users have certain attitudes and show certain behavior that cannot be captured in a numerical score on a scale [32]. A numerical score may highlight an issue as was demonstrated in the eeriness rating comparisons in Studies 1 and 2. However, a numerical score alone is unlikely to provide guidance on which aspects impact the reported attitudes and the subtleties in participants' reasonings when answering scale items. This understanding is especially important for semi-autonomous and embodied systems such as robots, which can invoke complex mental models and social-affective reactions. Researchers have begun to collect qualitative information about people's attitudes towards telepresence robots using ethnographic techniques (e.g., [33]).

The work presented in this paper is different than almost all of the research conducted so far with NARS (Table II) because telepresence robots inherently have a person in the loop. Overall, the results from our three studies indicated that NARS may be applied to telepresence robots. Additional work must be done to conclusively validate these scales for use with interactions with telepresence robots from the perspective of people physically present with the robot. Since the field of telepresence robots is in its infancy, now is the right time to figure out what people's attitudes and anxieties towards these robots are and take corrective action, if needed.

Acknowledgments

This research was funded in part by the National Science Foundation (IIS-0546309, IIS-0905228) at the University of Massachusetts Lowell, the ICIS project (BSIK03024) at the University of Amsterdam by the Dutch Ministry of Economical Affairs, and the LIREC project funded by the European Community's Seventh Framework Programme (FP7/2007-2013) under grant agreement no. 215554.

The authors would like to thank Dr. Chris Uhlik of Google Engineering Research for hosting the Study 2 experiment as well as Anybots and VGo Communications for the use of their beta stage robots. The authors would like to thank Dan Brooks of UMass Lowell for his assistance in defining Study 3 and VGo Communications for the loan of a robot. The authors would also like to acknowledge Alia Amin of CWI, Vanessa Evers of the University of Amsterdam, and Bob Wielinga of the University of Amsterdam for their contributions to the Robosapien study. We thank Del Siegle of the University of Connecticut for use of his reliability calculator [34]. Finally, we thank the reviewers for their comments and suggestions for improving this article.

References

[1] T. Nomura, T. Suzuki, T. Kanda, and K. Kato, "Measurement of Negative Attitudes toward Robots," *Interaction Studies*, vol. 7, no. 3, pp. 437–454, 2006.

[2] T. Nomura, T. Kanda, T. Suzuki, and K. Kato, "People's Assumptions about Robots: Investigation of their Relationships with Attitudes and Emotions toward Robots," in *Proc. of IEEE Intl. Workshop on Robots and Human Interactive Communication*, 2005, pp. 125–130.

[3] T. Nomura, T. Suzuki, T. Kanda, and K. Kato, "Measurement of Anxiety toward Robots," in *Proc. of the 15th IEEE Intl. Symp. on Robot and Human Interactive Communication*, 2006, pp. 372–377.

[4] J. Kelley, "An Iterative Design Methodology for User-Friendly Natural Language Office Information Applications," *ACM Transactions on Information Systems (TOIS)*, vol. 2, no. 1, pp. 26–41, 1984.

[5] T. Nomura, T. Kanda, T. Suzuki, and K. Kato, "Psychology in Human-Robot Communication: An Attempt through Investigation of Negative Attitudes and Anxiety toward Robots," in *Proc. of the 13th IEEE Intl. Workshop on Robot and Human Interactive Communication*, 2004, pp. 35–40.

[6] T. Nomura and T. Kanda, "On Proposing the Concept of Robot Anxiety and Considering Measurement of It," in *Proc. of the 12th IEEE Intl. Symp. on Robot and Human Interactive Communication*, 2003, pp. 373–378.

[7] T. Nomura, T. Kanda, and T. Suzuki, "Experimental Investigation into Influence of Negative Attitudes toward Robots on Human-Robot Interaction," *AI & Society*, vol. 20, no. 2, pp. 138–150, 2006.

[8] C. Bartneck, T. Suzuki, T. Kanda, and T. Nomura, "The Influence of People's Culture and Prior Experiences with Aibo on Their Attitude towards Robots," *AI & Society*, vol. 21, no. 1, pp. 217–230, 2007.

[9] T. Nomura, T. Kanda, T. Suzuki, and K. Kato, "Prediction of Human Behavior in Human-Robot Interaction Using Psychological Scales for Anxiety and Negative Attitudes toward Robots," *IEEE Transactions on Robotics*, vol. 24, no. 2, pp. 442–451, 2008.

[10] C. Bartneck, T. Nomura, T. Kanda, T. Suzuki, and K. Kennsuke, "A Cross-Cultural Study on Attitudes towards Robots," in *Proc. of the Intl. Conf. on Human-Computer Interaction*, 2005.

[11] K. Tsui, M. Desai, H. Yanco, H. Cramer, and N. Kemper, "Using the "Negative Attitude Toward Robots Scale" with Telepresence Robots," in *Proc. of the Performance Metrics for Intelligent Systems Workshop (PerMIS)*, 2010.

[12] T. Nomura, T. Shintani, K. Fujii, and K. Hokabe, "Experimental Investigation of Relationships between Anxiety, Negative Attitudes, and Allowable Distance of Robots," in *Proc. of the 2nd IASTED Intl. Conf. on Human Computer Interaction*. ACTA Press, 2007, pp. 13–18.

[13] T. Nomura, T. Suzuki, T. Kanda, and K. Kato, "Altered Attitudes of People toward Robots: Investigation through the Negative Attitudes toward Robots Scale," in *Proc. AAAI Workshop on Human Implications of Human-Robot Interaction*, 2006, pp. 29–35.

[14] H. Cramer, A. Amin, V. Evers, and N. Kemper, "Touched by Robots: Effects of Physical Contact and Proactiveness," in *Proc. of the ACM SIGCHI Conf. Workshop on "The Reign of Catz and Dogz"*, 2009.

[15] H. Cramer, N. Kemper, A. Amin, B. Wielinga, and V. Evers, "'Give Me a Hug': The Effects of Touch and Autonomy on People's Responses to Embodied Social Agents," *Computer Animation and Virtual Worlds*, vol. 20, no. 2-3, pp. 437–445, 2009.

[16] D. Syrdal, K. Dautenhahn, K. Koay, and M. Walters, "The Negative Attitudes towards Robots Scale and Reactions to Robot Behaviour in a Live Human-Robot Interaction Study," in *Proc. of the AISB Symp. on New Frontiers in Human-Robot Interaction*, 2009.

[17] L. Riek, T. Rabinowitch, P. Bremner, A. Pipe, M. Fraser, and P. Robinson, "Cooperative Gestures: Effective Signaling for Humanoid Robots," in *Proc. of the* 5^{th} *ACM/IEEE Intl. Conf. on Human-Robot Interaction*, 2010, pp. 61–68.
[18] Amazon, "Amazon Mechanical Turk: Artificial Artificial Intelligence," 2009, http://www.mturk.com. Accessed Dec. 2009.
[19] C. Bartneck, K. Takayuki, H. Ishiguro, and N. Hagita, "My Robotic Doppelgänger–A Critical Look at the Uncanny Valley," in *Proc. of the* 18^{th} *IEEE Intl. Symp. on Robot and Human Interactive Communication*, 2009.
[20] A. Powers and S. Kiesler, "The Advisor Robot: Tracing People's Mental Model from a Robot's Physical Attributes," in *Proc. of the 1st ACM SIGCHI/SIGART Conf. on Human-Robot Interaction*, 2006, pp. 218–225.
[21] J. Monahan, "I Don't Know It But I Like You The Influence of Non-conscious Affect on Person Perception," *Human Communication Research*, vol. 24, no. 4, pp. 480–500, 1998.
[22] D. George and P. Mallery, *SPSS for Windows Step by Step: A Simple Guide and Reference, 11.0 update*. Allyn & Bacon, 2003.
[23] Anybots, "Your Personal Avatar," 2010, http://anybots.com. Accessed Dec. 2010.
[24] VGo Communications, "Introducing VGo. From Anywhere. Go Anywhere." 2010, http://vgocom.com. Accessed Dec. 2010.
[25] L. Takayama, V. Groom, and C. Nass, "I'm Sorry, Dave: I'm Afraid I Won't Do That: Social Aspects of Human-Agent Conflict," in *Proc. of the* 27^{th} *Intl. Conf. on Human Factors in Computing Systems*, 2009, pp. 2099–2108.
[26] S. Adjalgeirsson, "MeBot: A Robotic Platform for Socially Embodied Telepresence," Master's thesis, Massachusetts Institute of Technology, 2009.
[27] Peak Survival, "Desert Survival: Survival Tips, Survival Guide," 2010, http://www.peaksurvival.us/Desert_Survival.html. Accessed Dec. 2010.
[28] M. Minsky, "Telepresence," *Omni*, vol. 2, no. 9, pp. 45–52, 1980.
[29] J. Markoff, "The Boss Is Robotic, and Rolling Up Behind You," The New York Times, Sept. 2010, http://www.nytimes.com/2010/09/05/science/05robots.html?_r=1, accessed Sept. 2010.
[30] J. Sutter, "The Robot that Visits your Cubicle," CNN Tech, Jul. 2010, http://articles.cnn.com/2010-07-30/tech/anybots.robots.office_1_consumer-robots-blackwell-conversation?_s=PM:TECH, accessed Feb. 2011.
[31] NBC, "Robot Helps 'Bubble Boy' Make the Grade," Today Show, Feb. 2011, http://today.msnbc.msn.com/id/26184891/vp/41641714#41641714, accessed Feb. 2011.
[32] K. Hööök, "User-Centred Design and Evaluation of Affective Interfaces," in *From Brows to Trust*, ser. Human-Computer Interaction Series, Z. Ruttkay and C. Pelachaud, Eds. Springer Netherlands, 2005, vol. 7, pp. 127–160.
[33] M. K. Lee and L. Takayama, "'Now, I have a body': Uses and social norms for mobile remote presence in the workplace," in *Proc. of CHI*, 2011, to appear.
[34] D. Siegle, "Reliability Calculator," 2000, available at http://www.gifted.uconn.edu/siegle/research/Instrument%20Reliability%20and%20Validity/Reliability.htm. Accessed Aug. 2010.
[35] L. Cronbach, "Coefficient Alpha and the Internal Structure of Tests," *Psychometrika*, vol. 16, no. 3, pp. 297–334, 1951.
[36] R. Likert, "A Technique for the Measurement of Attitudes," *Archives of Psychology*, vol. 22, no. 140, p. 55, 1932.
[37] J. Snider and C. Osgood, *Semantic Differential Technique: A Sourcebook*. Aldine Pub. Co., 1969.
[38] J. Gliem and R. Gliem, "Calculating, Interpreting, and Reporting Cronbach's Alpha Reliability Coefficient for Likert-Type Scales," in *Proc. of Midwest Research to Practice Conference in Adult, Continuing, and Community Education*, 2003, pp. 82–88.
[39] R. Devellis, *Scale Development*. Sage Publications, 1991.
[40] D. P. Nichols, "My Coefficient a is Negative!" 1999, http://www.ats.ucla.edu/stat/spss/library/negalpha.htm. Accessed Dec. 2010.
[41] J. Cortina, "What is Coefficient Alpha? An Examination of Theory and Applications," *J. of Applied Psychology*, vol. 78, pp. 98–98, 1993.
[42] J. Nunnaly, *Psychometric Theory*. New York: McGraw-Hill, 1978.

Appendix

A.1. Establishing Consistency

Data collected from participants using self-reporting methods may have issues with consistency. People may be unaware of exactly how they feel when provided too fine a granularity for scale questions. For example, a person may be able to identify how happy he or she feels when provided with a 5- or 7-point scale. However, he or she may be unable to identify their precise choice on a 100-point scale; what exactly is the difference between a happiness level of 66 points versus 67 points? Also, when orally proctored, people may under-report on negatively phrased scale questions or over-report on positively phrased scale questions. One way of ensuring that a person is consistent with his/her responses is to have several related questions. Surveys could ask a question in a positive manner, then repeat the questions with a negative phrasing, or have redundant questions with entirely different phrasing which focus on the same dimension.

Cronbach's alpha coefficient measures the internal consistency of related questions [35]. Gliem and Gliem describe how to compute the Cronbach alpha value for a Likert scale [36] (e.g., 1 through n value scale anchored by "I strongly agree" to "I strongly disagree") and differential scale [37] (e.g., 1 through m value scale anchored by "hard" and "easy") questions in [38]. The formula for Cronbach's alpha from [39], [40] is

$$\alpha = \frac{K}{(K-1)}[1 - \frac{\sum_{i=1}^{K} \sigma_{Y_i}^2}{\sigma_x^2}] \quad (1)$$

where K is defined as the number of related questions and σ_x^2 is the overall scale variance. $\sum_{i=1}^{K} \sigma_{Y_i}^2$ is the sum of variance of the i^{th} question of the sample over all K questions. Cortina [41] discusses other definitions of the alpha value used to measure scale consistency.

The Cronbach alpha value can range from $-\infty$ to 1, although only values greater than 0 are meaningful [40]. Nunnaly established that an alpha value of 0.7 or higher is a sufficient reliability coefficient [42]. George and Mallory categorize alpha values of 0.90-1.00 as "excellent," 0.80-0.89 as "good," 0.70-0.79 as "acceptable," 0.60-0.69 as "questionable," 0.50-0.59 as "poor," and 0.00-0.49 as "unacceptable" [22]. Gliem and Gliem provide a reminder that the Cronbach alpha value does not ensure consistency for single item groupings [38] because when K=1, a divide by zero error will occur.

Katherine M. Tsui is a PhD candidate in the Computer Science Department at the University of Massachusetts Lowell. She also has a BS, MS, and Human-Computer Interaction certificate from UMass Lowell. Her research is at the intersection of human-robot interaction and assisitive technology and has included projects such as indirect human-robot interaction for bystanders and visual control of a wheelchair mounted robot arm for people cognitive impairments. The aim of her dissertation research is to develop a user interface and autonomous robot behaviors for a telepresence robot to be used by people with disabilities who reside in medical institutions.

TABLE VIII
STUDY 3: DESCRIPTION AND PRO AND CON POINTS FOR EACH CHOICE IN A MODIFIED DESERT SURVIVAL SCENARIO

Item	Description	Pros	Cons
Two bananas	Each banana has: 200 calories, 51 grams of carbs (28 g sugar), 2 g protein.	Carbs provide a burst of short term energy.	May spoil in heat if not eaten quickly.
OR			
One packet of peanuts	1 oz bag salted peanuts has: 150 calories, 11 grams of carbs (4 g sugar), 4 g protein.	Protein provides longer term energy.	Salt may cause thirst.
Emergency car blanket	Silver; 8 ft x 10 ft; weighs 1 lb.	Provide warmth during night. If sunny, more visible than the parachute for signaling. Can be used to extract water via solar still.	Does not provide shelter during the day. Building a solar still requires more energy than gained.
OR			
Red and white parachute	Nylon; 26 ft wide canopy; weighs 15 lbs.	Shelter during the day. If cloudy, more visible than emergency blanket for signaling. Comes with rope.	Nylon draws warmth from your body at night.
Knife	Rusty machete; 18 inches; no sheath	Can be used to rig shelter, cut cactus for water, used as a hammer. Generally useful.	Can hurt self. Cannot use as signal.
OR			
Pistol	2 shots preloaded.	Can be used to signal or hunt. Gun powder can be used to start a fire. Use as a hammer.	Limited bullets. If sand gets in, the pistol won't fire and therefore can't be used as a signal.
Map and compass	Topological map with town marking and digital compass.	Can be used for navigation to move towards town.	Limited food. Cannot create heat. Cannot signal.
OR			
Matches and book	50 waterproof matches and a book of poisonous of plants and animals.	Can create fire for warmth at night, signaling, and cooking animals.	Must remain stationary and wait for rescue

Munjal Desai is a PhD candidate in the Computer Science Department at the University of Massachusetts Lowell. He has a MS in Computer Science from the University of Massachusetts Lowell and a BE in Information Technology from Bombay University. His research interests include adjustable autonomy for robots, influence of trust in controlling autonomous robots, designing intuitive interfaces, and human-robot interaction in general. His Master's thesis focused on creating a sliding scale autonomy system that provided fine grained control of the robot. As part of his doctoral dissertation, he is working to model control allocation and trust for autonomous robots in the urban search and rescue (USAR) domain.

Holly A. Yanco is a Professor in the Computer Science Department at the University of Massachusetts Lowell. She has a PhD and MS in Computer Science from the Massachusetts Institute of Technology (MIT) and a BA in Computer Science and Philosophy from Wellesley College. Her research interests include human-robot interaction, multi-touch computing, interface design, robot autonomy, fostering trust of autonomous systems, and evaluation methods. Application domains for this research include assistive technology and urban search and rescue (USAR). Dr. Yanco is the General Chair of the 2012 ACM/IEEE International Conference on Human-Robot Interaction. She served on the Executive Council of the Association for the Advancement of Artificial Intelligence (AAAI) from 2006-2009.

Henriette Cramer is an postdoctoral researcher at the Swedish Institute of Computer Science (SICS) and the Mobile Life Centre in Stockholm, Sweden. She has a PhD from the University of Amsterdam, focusing on people's reactions to adaptive and autonomous systems, including robots. She works in the areas of mobile technology and robots, including their mediation in human-human contact. Her current research focuses on user studies and concept development in the area of mobile experiences, location-based services, wide distribution of research apps, mobile mash-ups and interaction with autonomous systems, robots and "things."

Nicander Kemper is currently a software testing specialist for KZA b.v. (Baarn, The Netherlands). KZA is specialized in software testing and quality assurance. Nicander received a M.Sc. degree in Information Sciences at the Human Computer Studies Lab at the University of Amsterdam. His research focused on Human Robot Interaction (HRI) under the supervision of Dr. Vanessa Evers. The goal of his work in general is to find ways for people to interact with complex robots or computer systems in a more intuitive and natural way. In particular he studied the effects of a robot touching a human combined with the robot's proactive behavior.