Cooperative Gestures: Effective Signaling for Humanoid Robots

Laurel D. Riek*, Tal-Chen Rabinowitch†, Paul Bremner‡, Anthony G. Pipe‡, Mike Fraser‡, and Peter Robinson*

*Computer Laboratory, University of Cambridge, UK
†Centre for Music and Science, University of Cambridge, UK
‡Bristol Robotics Laboratory, UK
Correspondence Email: Laurel.Riek@cl.cam.ac.uk

Abstract—Cooperative gestures are a key aspect of human-human pro-social interaction. Thus, it is reasonable to expect that endowing humanoid robots with the ability to use such gestures when interacting with humans would be useful. However, while people are used to responding to such gestures expressed by other humans, it is unclear how they might react to a robot making them. To explore this topic, we conducted a within-subjects, video-based laboratory experiment, measuring time to cooperate with a humanoid robot making interactional gestures. We manipulated the gesture type (beckon, give, shake hands), the gesture style (smooth, abrupt), and the gesture orientation (front, side). We also employed two measures of individual differences: negative attitudes toward robots (NARS) and human gesture decoding ability (DANVA2-POS). Our results show that people cooperate with abrupt gestures more quickly than smooth ones and front-oriented gestures more quickly than those made to the side, people's speed at decoding robot gestures is correlated with their ability to decode human gestures, and negative attitudes toward robots is strongly correlated with a decreased ability in decoding human gestures.

Index Terms—affective robotics; cooperation; gestures; human-robot interaction

I. INTRODUCTION

In human-human collaboration, cooperative gestures play a key role in helping communicate intent, instruct, lead, and build rapport. From as early as 14 months of age, humans communicate cooperatively, to inform others of things and to share interest [1]. Nehaniv [2] refers to these gestures as *interactional*, used to "regulate interaction with a partner, i.e. used to initiate, maintain, invite, synchronize, organize or terminate a particular interactive, cooperative behavior." Thus, it seems reasonable to expect that affording such gestural interaction to people collaborating with robots might be helpful.

To date much work has been done on building robots capable of recognizing human gestural cues [3]–[5]; however, an equally important area to successful human-robot collaboration is making robots capable of generating meaningful, recognizable gestural cues to humans. Mutlu et. al [6] and Staudte and Crocker [7] have worked on addressing this problem for gaze while others have looked at effective generation of robot facial expressions (see Breazeal et. al [8] for a recent survey). For perception of expressive robot arm and body gestures, Kozima et al. [9] looked at children's overall perceptions of movements, Kim et al. [10] explored how varying gesture expressiveness can affect people's perceptions of a robot's personality, and Kulic and Croft [11] studied how to use

Fig. 1. A sample of interactional gestures used in human communication. *Photo credit (from top left): Jan Koch, Neil Mallett, I, Timmy, and Dan Zen.*

implicit measures to best alter a robot arm's movement in real time to put people at ease.

In our previous work, we sought to understand how people could empathize with robots of varying degrees of human-likeness [12], [13]. However, as we discovered during the course of our research, this problem is multi-faceted. In addition to the physical appearance of the robot, a plethora of aspects about its behavior may also play a role in how humans empathize with it, such as its speech, gaze, gestures, posture, etc. Ultimately, these are all aspects of a robot's *expressivity*, and for this work we decided to focus specifically on people's perception of and cooperation with a humanoid robot expressing itself via interactional arm gestures.

When humans cooperate with one another, they are typically well-adept at inferring another's intentions and emotions in order to spontaneously coordinate to achieve shared goals [14]. This cooperative ability is the pro-social, behavioral component of empathy, thought to lie at the core of "pure" empathy [15], [16]. It is rapid and comes naturally to us when we are interacting with other humans who communicate with us using interactional gestures. However, it is unclear how this ability might be affected when interacting with humanoids.

Thus, we are interested in several questions. First, how does the *style* of gesture affect people's cooperation with it? For example, is it better for humanoid robots to be more "smooth"
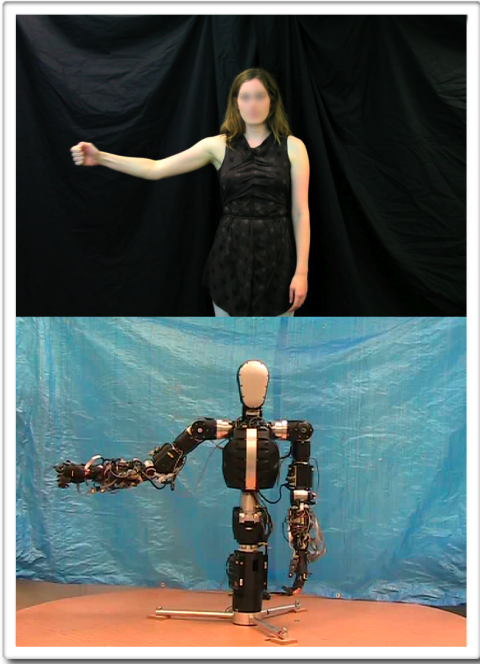
Fig. 2. A still frame from one of the human gesture videos used to instruct the robot programmer and the corresponding frame from the stimuli video.

and human-like in the way they form gestures, or should they be "abrupt" and machine-like? Work by Goetz et al. [17] indicates that it is important that a robot's appearance matches people's expectations, and people might expect a humanoid robot to make gestures in a human-like manner. Furthermore, Kanda et al. [18] showed that people reveal subtle, inherent biases in their nonverbal behaviors favoring more human-like robots.

Our second question is how much *precision* is necessary for a gesture to convey intention? If a robot gestures directly toward a human, or off to the side, does that make a difference? The reason we are interested in this question is one of practicality - not all robots have the necessary effectors and degrees-of-freedom to gesticulate clearly. However, it may be that people's skill in understanding gestures would not be inhibited by imprecision, thus affording even the most basic robots a means for communicating primitive social signals [19].

Finally, our third question is: regardless of the style or precision a robot uses when gesturing, do individual differences in interpersonal sensitivity impact peoples' cooperation with the robot? In human-human interaction people's general ability to encode and decode verbal and non-verbal behavior is extremely important in daily life, and some people are naturally more skilled at it than others [20]. Might we find a similar variability among people faced with decoding robot gestures and responding to them?

We designed an experiment to address these questions. In the experiment, we gave participants an interaction scenario that involved them watching video of a humanoid robot making various interactional gestures and being asked to take an action in response. We manipulated the orientation, style, and type of gesture the robot used. We measured participants' response time during this interaction, and used this as our primary measure of cooperation.

We expected people would be faster to cooperate with gestures made to the front as opposed to the side because front-oriented gestures are more commonly used in daily life so people are more used to them. Secondly, we expected that people would be faster to cooperate with gestures made in a smooth style as opposed to an abrupt one, because the humanoid robot used in our experiment was capable of enacting very human-like motion. Finally, we expected that people who are better at decoding human gestures would also be faster to respond to robot gestures. Our results, described in Section IV, support our first and third expectations, but negate our second. These results are discussed in detail in Section V.

II. MATERIALS

A. BERTI Robot

In our experiment we used our bespoke humanoid robot torso BERTI (Bristol and Elumotion, Robotic Torso 1.) (See Fig. 2, bottom.) Each arm of the robot has seven degrees of freedom (DOFs), each hand has nine DOFs, and the neck and waist have two DOFs. This gives a total of 36 DOFs to the robot. The joints in the arm are able to achieve speeds similar to those of human movement. The finger and wrist joints, while not capable of replicating human speeds because of mechanical limitations, are able to move in a way that is adequate for human-like gesturing. Thus, BERTI is able to make human-like gestures [21].

B. Stimulus Videos

After studying the literature and consulting with some gesture experts, we selected three interactional gestures to use in our experiment. The gestures are described in detail in Table I, but briefly they are: Beckon (the robot indicates the viewer should follow), Give (the robot pantomimes giving something to the viewer), Shake Hands (the robot indicates it would like to shake hands with the viewer.)

Human gestures typically consist of three phases: preparation, stroke, and retraction. First, the subject moves their hand into position, then the subject makes the gesture, and finally the subject's hands return to the resting position [22]. The gestures used in our experiment also followed this pattern. (See Fig. 4 for an example).

In order to give precise instructions to the robot programmer, we first filmed a human making each of the three aforementioned gestures (Beckon, Give, and Shake Hands), across two orientations (front and side), and in two different styles (smooth and abrupt). This made for a total of 12 videos. The videos were filmed against a dark background, and the camera was a fixed distance away from the subject. All videos were filmed from the torso up, with the subject facing the camera directly at all times, even when gesturing to the side. The right arm was the only arm used for gesturing - the left arm remained still at the side. (See Fig. 2, top).

TABLE I
THE THREE INTERACTIONAL GESTURES USED IN THE EXPERIMENT

Gesture Name	Action Response Label	Description
Beckon	Follow	The robot raises its arm, palm upward, towards its face, makes two strokes, then retracts its arm down.
Give	Take	The robot raises a closed fist to the center of its torso, then extends its hand, then retracts its arm down.
Shake Hands	Shake Hands	The robot moves its right arm, palm facing to the left up, and then fully forward, then retracts its arm down.

The 12 robot videos were filmed in exactly the same way as the human videos were filmed, though we were able to exert much more precise control over the speed of the robot's movements. This was particularly important for our style manipulation, as the speed of gesture can significantly affect its perception. Thus, for the "smooth" gestures the average speed of the robot was kept constant across the three gestural phases (preparation, stroke, retraction), but for the "abrupt" gesture the average speed was executed five times faster during the stroke phase. See Fig. 3 for an example illustration of this.

There are a wide range of models of human arm motion, such as the minimum jerk model [23] and the minimum commanded torque change model [24], to simplify motion production while still producing demonstrably well rated gestures [21]. For programming BERTI, we used a control scheme that uses key features present in all the models - smooth, direct motion between the end points. We define a smooth trajectory in this context as one where the acceleration profile of the movement contains no discontinuities. Thus, the "abrupt" motions are less human-like because the hand trajectory is less smooth, but without user-identifiable reasons for the disfluency.

In order to ensure all the stimuli were comparable, we set each video to have exactly five frames of stillness (with the robot's arms at its side) before the preparation phase began. We also made sure each video was exactly 8000 ms long by

Fig. 3. Velocity of the robot's hand during *Shake Hands*. The abrupt velocity profile shows the difference in acceleration used during the stroke phase. Otherwise the acceleration is constant between the two gesture styles.

duplicating the final post-retraction "at rest" frame of stillness the requisite number of times at the end. Setting the videos to this precise standardization was performed using the program MPEG Streamclip.

All stimulus videos used in the experiment were silent because we did not want to conflate the perception of the gestures with speech. Our experimental design was already sufficiently complex that we decided to not study speech perception at this time. While it is true that in human-human communication nearly all speech is accompanied by gestures [25], the converse is not necessarily the case. So et. al [26] ran a study where participants were asked to describe complex vignettes without speech (gesture alone). They found subjects were perfectly capable of doing this successfully, and indeed turned their gestures into a language-like structure. Hubbard et. al [27] found that although speech and gesture share common neural substrates, and people's perception of gestures are influenced by the presence of speech, given context people are still capable of perceiving speechless motion in a meaningful way. Thus, because in our experiment we gave participants context via a scenario that primed them for gesture-only communication, and also trained them on the three gestures that would be used, we believe it was reasonable for our experimental stimuli to be silent.

C. Action Response Labeling Pilot

Before running our experiment, it was necessary to first generate the category labels participants would use as their action responses in the main experiment. To do this, we ran a labeling pilot study with unschooled (naïve) participants from the same sampling population as our primary experiment. This labeling approach has been successfully applied by other nonverbal behavior researchers [28] and seemed appropriate for our purposes. We used fixed choice labels following similar labeling experiments described in Hall and Bernieri [20].

Three participants took part in in our pilot study, and were paid for their participation. The participants' ages ranged from 24 to 28. One was female and two were male. Two were British nationals and one was American. Two participants were postgraduate students and one an administrator in local government. Participants were recruited in the same way as they were for our primary study (see Section III-A), but those who participated in the pilot did not participate in the main experiment.
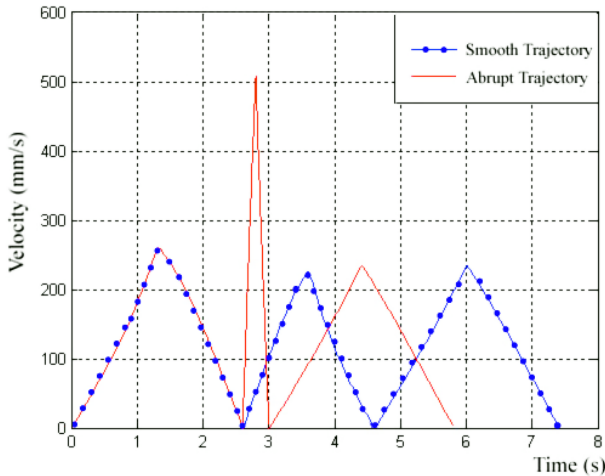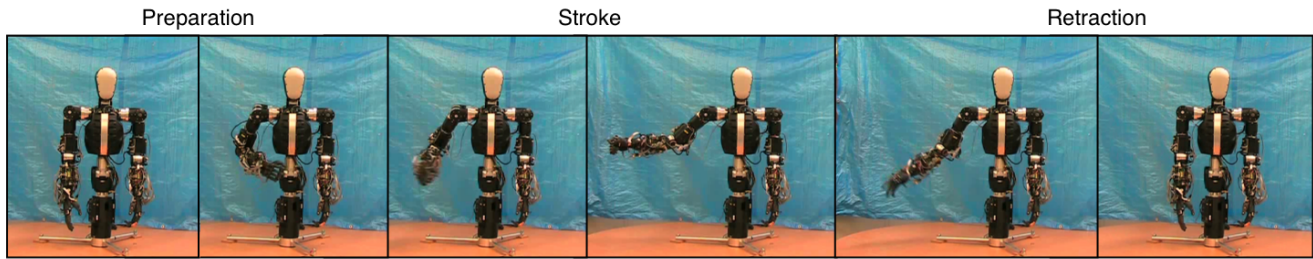
Fig. 4. An image sequence for the *Take* gesture made to the side, across the three gesture phases.

In the pilot, participants independently viewed each of the 12 stimulus robot videos and six of the human videos in random order. The six human videos were not used at all in the main experiment; we only included them in the pilot to get ground truth about how well we had programmed the robot. (Only the "smooth"-style version of the gestures needed to be labeled, because the "abrupt"-style gestures were nearly identical in terms of programming.)

For each video, participants were provided with three fixed choice labels, *Follow*, *Take*, and *Shake Hands*, and asked to choose what their action would be in response to what the robot was doing. To assess inter-rater reliability, we used Krippendorff's α, which is viewed as more reliable than other reliability measures when there are more than two raters [29]. Here, Krippendorff's α was 0.802, which indicates good reliability [30]. Thus, we feel confident in using these action labels in our main experiment.

D. Behavioral Measures

We prepared two post-experimental behavioral measures. The first was the Diagnostic Analysis of Nonverbal Accuracy for Postures (DANVA2-POS), a well-validated measure developed by Pitterman and Nowicki that assesses individual differences in ability to decode emotion in the body [31], [32]. The DANVA2-POS stimuli is a set of 40 still photographs depicting male and female participants in sitting or standing poses, using posture to convey one of four basic emotions (Happy, Sad, Angry, or Fearful) or neutral. All faces in the stimuli set are blacked out to ensure that facial expressions do not interfere with body posture interpretation.

We selected the DANVA2-POS instrument for our experiment because we wanted an objective measure of each participant's ability to decode nonverbal behavior in humans, to see if it had any bearing on their ability to decode a humanoid robot's gestures whilst interacting with it.

Our second post-experimental instrument was the Negative Attitudes Toward Robots Scale (NARS) developed by Nomura et al. [33]. This is a summed measure that assesses negative attitudes toward robots via a 5-point attitudinal scale. The measure contains three sub-scales: "negative attitudes toward emotions in interaction with robots," "negative attitudes toward the social influence of robots," and "negative attitudes toward situations of interaction with robots." We selected this instrument because recent validation work by Syrdal et al. [34]

suggests that the English version of this measure may be a reasonable means of assessing people's prior attitudes towards robots, attitudes which could influence how they might cooperate with a robot. The researchers also suggested that NARS is well-suited toward a British University population. Since we are primarily interested in how people interpret and respond to the style and manner in which a robot gestures toward them, and are sampling from a British University population, this seemed like a reasonable instrument to employ.

E. Scenario

In order to help participants engage in the experiment and put them at ease, we developed a realistic scenario about a fictitious robotics company. Participants were told that the "enFutura Corporation" developed a new series of robots designed to work with humans in various domestic settings; however, the robots are unable to communicate verbally and can only communicate via gestures. enFutura is unsure how to best program their robots, and the purpose of the experiment is to help provide them with feedback. This feedback is to be conveyed via playing a game, viewing videos featuring enFutura's robots, each one programmed slightly differently. In the game, the goal is to pick an action based on what the participant thinks the robot is asking him or her to do, as quickly as possible. Further, participants were told that when they pressed a key on the keypad, they were actually responding to the gesture the robot was making.

Following the experiment, participants were properly debriefed and told neither the scenario nor company were real.

F. Training Materials

We developed a training video for participants to view at the beginning of the experiment. The video was approximately a minute and a half long, and displayed two examples of each of the three gestures, displayed from the front and side, along with the action response labels (*Follow*, *Take*, *Shake Hands*) subtitled on each video. Before watching the video, participants were told that these labels do not identify the gestures the robot is making, but instead represent the actions they should take in response to the gestures.

We also prepared a short training session with the response pad that showed participants text of the three action labels, three times per label, presented in random order.

Fig. 5. The response time measurement device used in our experiment. *Image courtesy of Cedrus Corporation.*

G. Measurement Apparatus

The experiment was written and run in MATLAB version R2008b, using version PTB-3 of the Psychophysics Toolbox [35]. This package was used in order to ensure high precision when measuring reaction times. In particular for displaying our stimuli, we modified the *DetectionRTInVideoDemoOSX.m* program, which uses the OpenGL Renderer for OS X and takes into account the graphics card, processor speed, and VRAM when rendering videos in order to ensure highly precise reaction times. In particular, the machine which ran the experiment was a Macbook Pro, 2.2 Ghz Intel Core 2 Duo, with 2 GB of RAM and an NVIDIA GeForce 8600M with 128 MB of VRAM, running Mac OS X Version 10.5.7.

To capture participant reaction times accurately, we used the RB-530 Cedrus response pad (See Fig. 5). These response pads are frequently used in reaction time experiments as they typically have a higher degree of precision than keyboard or mouse inputs. Output from the pad was recorded using the Cedrus library of the Psychophysics Toolbox.

The response pad had text labels on the buttons containing the three action response labels. We also kept a paper card on the monitor throughout the experiment which showed a mock-up of the response pad, to serve as a reminder to participants during the experiment so they didn't have to look down.

III. METHODOLOGY

We conducted a 3 x 2 x 2, within-subjects, video-based experiment in the laboratory. We manipulated the gesture type (beckon, give, shake hands), the gesture style (smooth, abrupt), and the gesture orientation (front, side). This made for a total of 12 videos, each of which participants viewed four times, in random order across four trials.

A. Participants

Participants were recruited via University electronic bulletin boards and by word of mouth. They were told they would be taking place in a video-based experiment involving robot gestures, and would be paid a £5 gift voucher for their participation. We requested participants be aged 17 years or older, as the DANVA2-POS instrument is intended for adults.

A total of 16 participants participated in our experiment, 9 female and 7 male. Their ages ranged from 23 to 38, the mean age was 28.5, (s.d. = 4.07). In terms of nationality, six participants were British, two were Canadian, two were Western European (Portuguese, Austrian), two were South Asian (Indian, Bangladeshi), two were East Asian (Japanese, Chinese), and two were Southeast Asian (Singaporean). All participants considered themselves fluent in English. For the non-British participants, average time spent living in the UK was 2.33 years, (s.d. = 1.19). In terms of occupation, eight participants were current postgraduate students, one a recent MPhil graduate, two undergraduate students, one professional writer, one housewife, one project manager, one post-doc, and one unemployed person.

15 participants were right-handed and one was left-handed. Also, 15 participants considered themselves to have enjoyed a good night's sleep the night before the experiment and one did not. (Both of these factors can affect reaction time [36]).

B. Procedure

Upon arrival, participants received a sheet of written instructions detailing the overall flow of the experiment. Following this, they viewed the training video described in Section II-F which introduced the three gestures and their suggested response labels. Next, participants were given the interaction scenario described in Section II-E. Then, participants practiced using the response pad. The experimenter instructed participants to only use the index finger of their dominant hand when using the response pad to ensure uniformity across all trials of all participants and help reduce error.

At this point, the experimenter left the room, allowing the participant to complete the main experiment alone. The main experiment consisted of four trials, each trial consisting of the 12 videos presented in random order. After completing two trials participants had a short break by viewing a minute from a nature film.

After participants completed the main experiment, the experimenter joined them and asked them to complete a paper-based demographics sheet and the NARS measure. Following this, the experimenter manually administered the DANVA2-POS test using MATLAB, to control the precise presentation timings of the stimuli (two seconds each). Finally, participants were interviewed and debriefed by the experimenter, and paid for their participation.

C. Measures

Our dependent variables were cooperation speed, as measured via reaction time, and action response labeling accuracy, as measured against ground truth. Our independent variables were gesture type (beckon, give, shake hands), gesture orientation (front, side), and gesture style (smooth, abrupt). We also collected two behavioral measures, NARS and DANVA2-POS, which are summed-measure scores.

IV. RESULTS

A. Learning Effect Check

Before beginning our analysis, we first checked to see if there was a learning effect between trials. Even though the
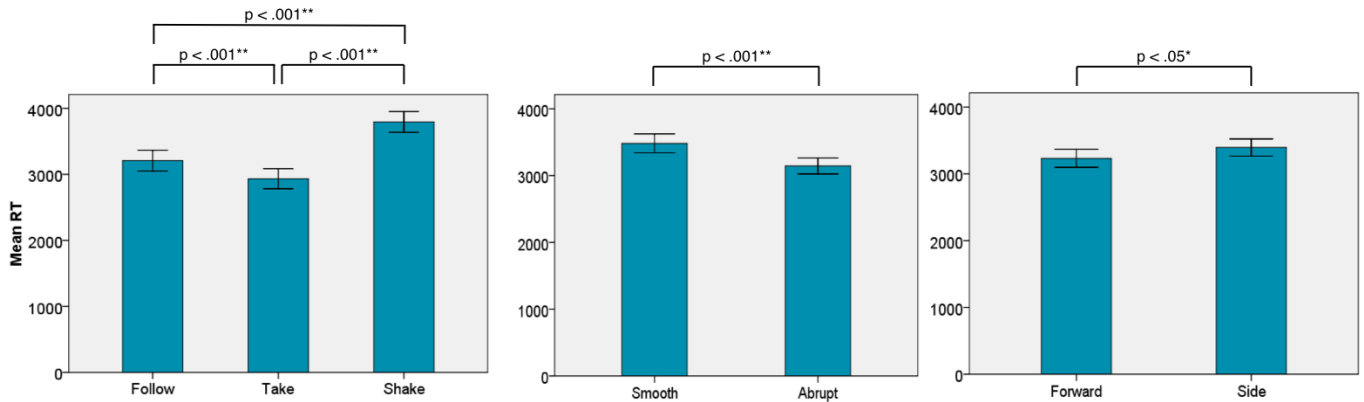
Fig. 6. Mean reaction times (across all trials) for each of the three manipulations: Gesture Type, Style, and Orientation. One asterisk (*) indicates statistically significant probabilities less than .05, two asterisks (**) less than .001.

presentation order of the videos was randomized within each of the four trials, we still wanted to ensure that trial order did not affect participants' labeling errors. In particular, we conducted a logistic regression to see if trial number (1-4) predicted action labeling errors. We define action labeling errors to be binary, and caused by participants either selecting a non-ground truth action label (e.g., the participant chose *Take* instead of *Follow*) or the participant chose no label at all (e.g., they timed out at 8000 ms).

We did not find a significant effect between the four trials, indicating that errors participants made are probably random, $b = -0.27$, $SE = 0.14$, $z = -1.87$, $p > .05$. Therefore, we will ignore labeling errors in our subsequent analysis.

Our error analysis did reveal one participant who contributed a disproportionate number of time-out errors to the total. Indeed, in the post-experimental interview the participant admitted that he "sometimes didn't feel like" cooperating with the robot even when he knew what the robot was asking him to do. Thus, we consider this participant to be an outlier and excluded their data from the analysis.

B. Missing Values

Across the entire data set, only three values were missing due to time-outs. To deal with this missing data, we tried both excluding the value (thus the entire trial for that participant) as well as three re-coding techniques: treating the value as an average, maximum, or minimum from the other three trials for that video. We found no significant impact on the F-values, p-values, or sphericity for our main ANOVA interactions. Thus, we will simply treat the each of the missing values as an average of its three neighboring trials.

C. Results

The following results refer to a three-way, repeated measures mixed ANOVA, with gesture type, orientation, and style as our within-subjects variables, and subject number as our between-subjects variable.

1) Gesture Type: There was a significant main effect of gesture type, $F(2, 90) = 79.50$, $p < .001$. Contrasts revealed that participants were faster to cooperate with the *Take* gesture compared with the *Follow* gesture, $F(1, 45) = 17.70$, $p <$

.001, $r = .53$, and also faster to cooperate with *Take* compared with *Shake Hands*, $F(1, 45) = 158.09$, $p < .001$, $r = .88$. Furthermore, participants were faster to cooperate with *Follow* compared with *Shake Hands*, $F(1, 45) = 62.79$, $p < .001$, $r = .76$. All reported contrasts are Bonferroni corrected.

2) Gesture Style: We found a significant main effect of gesture style. Participants were faster to cooperate with abrupt gestures compared with smooth ones, $F(1, 45) = 42.17$, $p < .001$, $r = .70$.

3) Gesture Orientation: There was a significant main effect of gesture orientation, participants were faster to cooperate with gestures made to the front than to the side, $F(1, 45) = 5.80$, $p < .05$, $r = .34$.

*4) Gesture Type * Gesture Orientation:* There was a significant interaction between gesture type and orientation, $F(2,90) = 17.60$, $p < .001$. This indicates that orientation had different effects on cooperation times depending on which gesture type was used. Contrasts revealed a significant interaction when comparing side-oriented gestures to forward-oriented gestures for *Take* compared with *Follow*, $F(1, 45) = 16.83$, $p < .001$, $r = .52$, and for *Follow* compared with *Shake Hands*, $F(1, 45) = 27.87$, $p < .001$, $r = .62$. (Contrasts are Tukey corrected.). The interaction graph shown in Fig. 7 reflects that side-oriented gestures increased reaction time significantly more in *Shake Hands* compared with *Follow*, and *Follow* compared with *Take*.

5) Individual Differences: DANVA2-POS scores ranged from 14-27, mean = 20.94, s.d. = 3.87. NARS scores ranged from 22-46, mean = 36.50, s.d. = 7.58.

A partial, two-tailed correlation controlling for participant number and video number revealed a significant correlation with a medium effect size between DANVA2-POS score and reaction time, $r = -.385$, $p < .001$. Thus, a higher DANVA2-POS score is associated with a quicker reaction time.

Removing two bivariate outliers based on their Cook's distance, we found a significant correlation with a large effect size between NARS and DANVA2-POS scores, $r = -.645$, $p < .05$. Thus, scores on the DANVA2-POS test are inversely correlated with NARS scores; people who are better at decoding human gestures hold more favorable views toward robots.
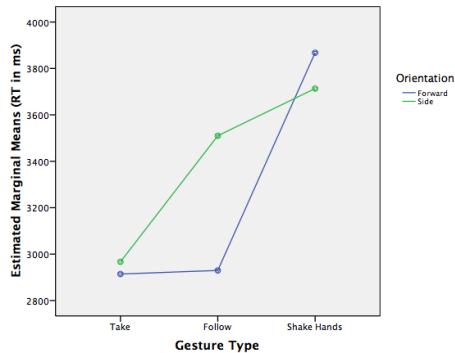
Fig. 7. The gesture type and orientation interaction. The y-axis represents the response time (RT) in milliseconds.

V. Discussion

Our findings suggest that there are several variables which significantly influence how people cooperate with robots while interacting with them. Our first finding was that gesture type significantly affected overall reaction time, regardless of style or orientation. Participants were significantly faster to cooperate with the *Take* gesture compared with *Follow*, and *Take* compared with *Shake Hands*, both with a large effect size. Also, participants were significantly faster to cooperate with *Follow* compared with *Shake Hands*, with a large effect size. From this study it is not possible to tell if these effects are unique to human-robot interaction, or whether they may also apply to human-human interaction. Further work is necessary in order to better understand the implications for designing gestures for humanoid robots.

Our second finding is that gesture style had a significant effect on cooperation time, regardless of gesture type or orientation. Participants were faster to respond to abrupt, non-human-like gestures as compared with smooth ones, with a large effect size. This result was surprising, because we expected the opposite due to the very human-like motion and human morphology BERTI afforded.

When measuring reaction time (RT), it is assumed that the time between stimulus and response is occupied by a sequence of processes and stages - some being mental operations, which occur one after the other [37]. Thus, a longer reaction time means a longer mental processing time. So for participants to have significantly longer reaction times for smooth gestures implies that abrupt gestures in robots require less mental processing. This may stem from a number of reasons. Since abrupt gestures are associated with robots more than with humans, it is possible that the discrepancy between the smooth gesture and the expectation of a robot to be abrupt expanded participants' mental processing time when deciding how to interact with the robot. Alternatively, it could very well be that the RT difference is associated with the way the movement is executed and consequently embodied in us, and processing abrupt movements may just prove to be faster than processing smooth ones. We cannot say for certain which of these two possibilities is correct, as this can only be determined by an additional human-human designed experiment.

Either way, the implication of this finding to the HRI community is the same. Abrupt gestures are communicated more quickly, and therefore more effectively than smooth gestures. In contexts where robots need to convey information as quickly as possible to humans, such as in safety-critical environments, it may behoove designers to consider programming robots to make more abrupt motions. As one participant mentioned, while these gestures can seem aggressive they do help to convey a sense of urgency.

Our third finding is that gesture orientation had a significant effect overall on cooperation time. As we expected, participants were faster to cooperate with gestures made to the front compared with those made to the side, with a medium effect size. This effect was most strongly seen for *Take* when compared with *Follow*, and for *Follow* compared with *Shake Hands*. Side-oriented gestures are rarely used in human-human communication for the interactional gesture types used in our experiment, so it is likely the slower reaction times reflects participants' needing additional mental processing to deal with the discrepancy. Again, for safety-critical environments where rapid communication is crucial, it would be good for robot designers to aim for as much orientation accuracy as possible when designing gestures for their robots. However, for robots intended to serve in companion, entertainment, or educational contexts, accurate orientation is not as crucial, because despite participants being slower to recognize side-oriented gestures, their response accuracy did not disproportionately suffer.

Finally, we found a significant relationship between DANVA2-POS scores and reaction times. This seems to imply that people who are better at decoding emotion in human gestures are also probably better at decoding robot gestures. We also found a highly significant relationship between DANVA2-POS scores and NARS scores, which means that, for our sample, people who have more negative attitudes toward robots also are less skilled at decoding human gestures. We find these results intriguing, and plan to see if they replicable in other experimental contexts.

A. Limitations

Ideally this study would be conducted with a live robot, because the literature is not entirely clear on whether using video-based stimuli is entirely equivalent to the use of a co-located robot. While work by Dautenhahn [38] suggests that video-based trials are a viable methodological approach and in some cases comparable to live HRI trials, work by Powers et al. [39] suggests that a projected robot compared with a co-located robot can cause differences in interaction.

In the future, we hope to use measures in addition to response time to further understand what it means to cooperate with robots via gestural interaction to help develop more concrete design guidelines. As previously mentioned, faster responses may be more important in some contexts than others. Thus, additional experiments would involve realistic scenarios across multiple contexts.

While we employed speechless gesture in this experiment for reasons described in Section II-B, doing so may have con-

flicted with participants' expectations per social convention. In interviews, several participants mentioned that normally one would use speech along with or proceeding the use of gestures to avoid coming across as rude. This idea was frequently mentioned for the *Follow* gesture, as this was seen by some as more of a command as opposed to a request.

Gestures are very much dependent on culture. While we had a diverse cultural sample of participants, they have all lived in the UK for a long time and have no doubt grown accustomed to British gestures. Our post-experimental interviews with participants suggested that the way these gestures are formed can greatly vary between cultures and by context. For example, in some East Asian cultures a robot servant that gave a person something with only one hand as opposed to two could be seen as rude. However, as one participant eloquently put it, "Context is more important than content." With adequate training, most people would be able to learn the meaning of particular gestures the robot made and would probably be able to tolerate any cultural faux pas.

REFERENCES

[1] H. Moll and M. Tomasello, "Cooperation and human cognition: the vygotskian intelligence hypothesis," *Philosophical Transactions of the Royal Society B: Biological Sciences*, vol. 362, no. 1480, 2007.

[2] C. Nehaniv, K. Dautenhahn, J. Kubacki, M. Haegele, C. Parlitz, and R. Alami, "A methodological approach relating the classification of gesture to identification of human intent in the context of human-robot interaction," in *Proc. of the IEEE Int'l Symposium on Robot and Human Interactive Communication (RO-MAN)*, Aug. 2005.

[3] C. Breazeal, G. Hoffman, and A. Lockerd, "Teaching and working with robots as a collaboration," in *Proc. of the Third Int'l Conference on Autonomous Agents and Multiagent Systems (AAMAS)*. IEEE, 2004.

[4] M. M. Loper, N. P. Koenig, S. H. Chernova, C. V. Jones, and O. C. Jenkins, "Mobile human-robot teaming with environmental tolerance," in *HRI '09: Proc. of the 4th ACM/IEEE Int'l conference on Human robot interaction*. 2009.

[5] C. L. Sidner, C. Lee, C. D. Kidd, N. Lesh, and C. Rich, "Explorations in engagement for humans and robots," *Artif. Intell.*, vol. 166, 2005.

[6] B. Mutlu, T. Shiwa, T. Kanda, H. Ishiguro, and N. Hagita, "Footing in human-robot conversations: how robots might shape participant roles using gaze cues," in *HRI '09: Proc. of the 4th ACM/IEEE international conference on Human robot interaction*. 2009.

[7] M. Staudte and M. W. Crocker, "Visual attention in spoken human-robot interaction," in *HRI '09: Proc. of the 4th ACM/IEEE international conference on Human robot interaction*. 2009.

[8] C. Breazeal, A. Takanishi, and T. Kobayashi, "Social robots that interact with people," in *Springer Handbook of Robotics*, B. Siciliano and O. Khatib, Eds. Springer, 2008, pp. 1349–1369.

[9] H. Kozima, M. Michalowski, and C. Nakagawa, "Keepon: A playful robot for research, therapy, and entertainment," *Int J Soc Robotics*, 2008.

[10] H. Kim, S. S. Kwak, and M. Kim, "Personality design of sociable robots by control of gesture design factors," in *The 17th IEEE Symposium on Robot and Human Interactive Communication, (RO-MAN)*, Aug. 2008.

[11] D. Kulic and E. A. Croft, "Affective state estimation for human–robot interaction," *IEEE Transactions on Robotics*, vol. 23, no. 5, 2007.

[12] L. D. Riek, T.-C. Rabinowitch, B. Chakrabarti, and P. Robinson, "Empathizing with robots: Fellow feeling along the anthropomorphic spectrum," in *Proc. Int'l Conf. on Affective Computing and Intelligent Interaction (ACII)*. Amsterdam, Netherlands: IEEE, 2009.

[13] L. D. Riek, T.-C. Rabinowitch, B. Chakrabarti, and P. Robinson, "How anthropomorphism affects empathy toward robots," in *HRI '09: Proc. of the 4th ACM/IEEE Int'l conference on Human robot interaction*. 2009.

[14] F. B. M. de Waal, *The 'Russian doll' model of empathy and imitation*. John Benjamins Publishing Company, 2007, ch. 3, pp. 49–69.

[15] S. D. Preston and F. B. M. de Waal, "Empathy: Its ultimate and proximate bases," *Behavioral and Brain Sciences*, vol. 25, no. 1, 2002.

[16] L. Wispe, *The psychology of sympathy*. New York: Plenum Press, 1991.

[17] J. Goetz, S. Kiesler, and A. Powers, "Matching robot appearance and behavior to tasks to improve human-robot cooperation," *The 12th IEEE Int'l Workshop on Robot and Human Interactive Communication (ROMAN)*, 2003.

[18] T. Kanda, T. Miyashita, T. Osada, Y. Haikawa, and H. Ishiguro, "Analysis of humanoid appearances in humanrobot interaction," *IEEE Transactions on Robotics*, vol. 24, no. 3, pp. 725–735, June 2008.

[19] A. Vinciarelli, M. Pantic, and H. Bourlard, "Social signal processing: Survey of an emerging domain," *Image Vision Comput*, vol. 27, 2009.

[20] J. A. Hall and F. J. Bernieri, Eds., *Interpersonal Sensitivity: Theory and Measurement*. Lawrence Erlbaum, 2001.

[21] P. Bremner, A. Pipe, M. Fraser, S. Subramanian, and C. Melhuish, "Conversational gestures in human-robot interaction," in *IEEE International Conference on Systems, Man, and Cybernetics*, 2009.

[22] A. Kendon, *Gesture: visible action as utterance*. Cambridge University Press, 2004.

[23] T. Flash and N. Hogan, "The coordination of arm movements: An experimentally confirmed mathematical model," *J Neurosci*, vol. 5, 1985.

[24] E. Nakano, H. Imamizu, R. Osu, Y. Uno, H. Gomi, T. Yoshioka, and M. Kawato, "Quantitative examinations of internal representations for arm trajectory planning: Minimum commanded torque change model," *Journal of Neurophysiology*, vol. 81, pp. 2140–2155, 1999.

[25] A. Kendon, "Gesticulation and speech: Two aspects of the process of utterance," in *The Relationship of Verbal and Nonverbal Communication*, M. Key, Ed. Mouton and Co., 1980, pp. 207–227.

[26] W. So, M. Coppola, V. Licciardello, and S. Goldin-Meadow, "The seeds of spatial grammar in the manual modality," *Cognitive Sci*, vol. 29, 2005.

[27] A. Hubbard, S. Wilson, D. Callan, and M. Dapretto, "Giving speech a hand: Gesture modulates activity in auditory cortex during speech perception," *Human brain mapping*, vol. 30, no. 3, 2009.

[28] R. J. Waldinger, M. S. Schulz, S. T. Hauser, J. P. Allen, and J. A. Crowell, "Reading others' emotions: The role of intuitive judgments in predicting marital satisfaction, quality, and stability." *Journal of Family Psychology*, vol. 18, no. 1, pp. 58–71, 2004.

[29] A. F. Hayes and K. Krippendorff, "Answering the call for a standard reliability measure for coding data," *Commun Methods Meas*, 2007.

[30] K. Krippendorff, *Content Analysis: An Introduction to Its Methodology*. Sage Publications, 2004.

[31] S. Nowicki and M. Duke, "Individual differences in the nonverbal communication of affect: The diagnostic analysis of nonverbal accuracy scale," *Journal of Nonverbal Behavior*, vol. 18, no. 1, pp. 9–35, 1994.

[32] H. Pitterman and S. Nowicki, "A test of the ability to identify emotion in human standing and sitting postures: The diagnostic analysis of nonverbal accuracy-2 posture test (danva2-pos)," *Genetic, Social, and General Psychology Monographs*, 2004.

[33] T. Nomura, T. Suzuki, T. Kanda, and K. Kato, "Measurement of negative attitudes toward robots," *Interaction Studies*, vol. 7, 2006.

[34] D. S. Syrdal, K. Dautenhahn, K. L. Koay, and M. L. Walters, "The negative attitudes towards robots scale and reactions to robot behaviour in a live human-robot interaction study," in *In Proc. of the AISB Symposium on New Frontiers in Human-Robot Interaction*, 2009.

[35] D. Brainard, "The psychophysics toolbox," *Spatial Vision*, vol. 10, 1997.

[36] R. J. Kosinski, "A literature review on reaction time," Clemson University, Tech. Rep., Sep. 2008. [Online]. Available: http://biae.clemson.edu/bpc/bp/Lab/110/reaction.htm

[37] S. Sternberg, "Memory-scanning: Mental processes revealed by reaction-time experiments," *Cognitive Psychology: Key Readings*, p. 48, 2004.

[38] K. Dautenhahn, "Methodology and themes of human-robot interaction: a growing research field." *Int J of Adv Robotic Sy*, vol. 4, no. 1, 2007.

[39] A. Powers, S. Kiesler, S. Fussell, and C. Torrey, "Comparing a computer agent with a humanoid robot," in *HRI '07: Proc. of the ACM/IEEE international conference on Human-robot interaction*. ACM, 2007.