How People Anthropomorphize Robots

Susan R. Fussell, Sara Kiesler, Leslie D. Setlock, Victoria Yew
Human Computer Interaction Institute
Carnegie Mellon University
5000 Forbes Avenue
Pittsburgh, PA 15206 USA
[sfussell, kiesler, lsetlock, vyew]@andrew.cmu.edu

ABSTRACT

We explored anthropomorphism in people's reactions to a robot in social context vs. their more considered judgments of robots in the abstract. Participants saw a photo and read transcripts from a health interview by a robot or human interviewer. For half of the participants, the interviewer was polite and for the other half, the interviewer was impolite. Participants then summarized the interactions in their own words and responded true or false to adjectives describing the interviewer. They later completed a post-task survey about whether a robot interviewer would possess moods, attitudes, and feelings. The results showed substantial anthropomorphism in participants' interview summaries and true-false responses, but minimal anthropomorphism in the abstract robot survey. Those who interacted with the robot interviewer tended to anthropomorphize more in the post-task survey, suggesting that as people interact more with robots, their abstract conceptions of them will become more anthropomorphic.

Categories and Subject Descriptors

H.1.2 [Models and Principles]: User/Machine Systems – *Human factors, Software psychology*. H.5.2 [Information Interfaces and Presentation]: User Interfaces – *Evaluation/methodology, Theory and methods*.

General Terms: Experimentation, Human Factors

Keywords: Human-robot interaction, social robots

1. INTRODUCTION

People's interactions with robots, other people, animals and physical objects are influenced by the ways in which they conceptualize these entities. Robots that can interact with people, unlike these other categories, are comparatively novel entities for which people may not have preconceptions. Robots that act socially and exhibit human-like appearance or behavior may be especially novel in the sense that people have few preconceptions about their underlying attributes and expected behaviors.

One way in which people can make sense of novel entities is by projecting existing social schemas onto them. Nass and his colleagues studied the imposition of social schemas onto computers and websites (e.g., [13]). A growing body of research suggests that people project social categories and social behaviors onto robots, particularly onto robots with humanoid characteristics or those engaged in social roles. They make assumptions about a robot's knowledge based on its surface features, including its gender features [20] and place of manufacture. [10]. People also make judgments about robots' personalities, based on their faces [30] or voice [14]. For example, baby-faced robots are thought to be more sociable than robots with other types of faces [21], and people project introversion and extroversion onto synthesized speech [15]. People also expect robots to show the kinds of social sensitivity found in human to human interaction, such as perspective-taking in communication [28].

Although clearly people attribute social properties to robots in research studies, it is less clear whether they believe that the robot literally possesses these characteristics (e.g., the robot *is* happy in the exact same way that a human is happy) or whether they instead are using human terms metaphorically (e.g., the robot is acting *as if* it were happy). It is also unclear whether people's immediate social responses to robots in context are the same as their more carefully considered judgments about robot properties in the abstract. People might act as if baby-faced robots are more sociable than robots with "adult" faces, for instance, but when asked explicitly about this belief, they might realize that there is no logical reason for this to be the case.

In the current experiment, we examine anthropomorphism in people's conceptualizations of robots at three levels of abstraction. We look at what they say about a robot's behavior in a specific context, at what properties they attribute to the robot performing these behaviors (e.g., it was frustrated, it was machine-like), and at what abstract properties they think characterize robots in general (e.g., robots can feel emotions, robots sometimes need repair). As we show in greater detail below, the results demonstrate considerable anthropomorphism in the communication and attribution tasks at levels 1 and 2, but little anthropomorphism in their level 3 responses about abstract robot properties.

1.1 Level 1: Free Description

People's implicit conceptualizations of robots may be revealed in the words they use to describe robot behavior. As Semin and Fiedler have shown in their Linguistic Category Model (e.g., [25], [26]), most any behavior can be described at different levels of abstraction. Consider, for example, the following four utterances:

(a) John *hit* Andrew
(b) John *abused* Andrew
(c) John *hated* Andrew
(d) John was *aggressive*

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

HRI'08, March 12–15, 2008, Amsterdam, Netherlands.
Copyright 2008 ACM 978-1-60558-017-3/08/03…$5.00.

The verb *hit* in utterance (a) is an example of a *descriptive action verb*—verbs that describe a single concrete event in objective, verifiable terms, such as *cry*, *walk* or *point*. These verbs describe what happened, but they do not impute intentions or goals to the actor. The verb *abused* in utterance (b) is an example of an *interpretive action verb.* Like descriptive action verbs, interpretive action verbs describe a single event but they are less verifiable. Observers might agree that John hit Andrew, but they may disagree as to whether this hitting constituted *abuse* as opposed to *self-defense*. Unlike descriptive action verbs, interpretive action verbs like *abused* also imply something about the performer of the action—in this case, that he or she is the type of person who would abuse others. The verb *hated* in utterance (c) is an example of a *state verb*—a verb that reflects a person's internal, nonverifiable experiences such as attitudes, emotions, or beliefs, Because they are removed from a particular instance of behavior, state verbs such as *feel*, *love* or *want* are more abstract than either descriptive or interpretive action verbs. Finally, adjectives such as *aggressive* in utterance (d) are the most removed from an actual instance of behavior and suggest traits that may generalize across many situations. An aggressive person may act aggressively with strangers and friends as well as with Andrew.

Choices among levels of abstraction when describing a robot's behavior reflect increasing levels of anthropomorphism. Consider the following modifications of the utterances above:

(e) The robot *hit* Andrew
(f) The robot *abused* Andrew
(g) The robot *hated* Andrew
(h) The robot was *aggressive*

As these examples show, the use of descriptive action verbs to refer to robot behavior is minimally anthropomorphic; mobile robots can in fact *hit* something and the verb is neutral as to whether or not the robot possessed an intention to hit Andrew. The use of interpretive action verbs to describe robot behavior is somewhat more anthropomorphic. *Abuse* implies something about the performer of the action, in this case, that the robot possessed the intention to cause harm to Andrew. State verbs like *hated* and adjectives like *aggressive* are even more anthropomorphic, because they impute internal cognitive states and emotions or personality traits to the robot.

1.2 Level 2: Attributions

We argue that it is comparatively easy for people to anthropomorphize a robot in conversation or free description, just by using ordinary words like "love" or "hate" casually. It takes a little more consideration and experience with a robot to answer a question about a robot when asked if it is "lovable" or "aggressive." Thus, a higher bar for anthropomorphism is raised when people are asked to make attributions of personality or character about a robot.

As people interact with other people, they spontaneously attribute personality characteristics, emotional states, and other social properties to them (e.g., [23]). When we hear that a woman is a nurse, for example, we might infer that she is caring, friendly and reliable. We therefore can expect anthropomorphism of robots in people's attribution processes—both their decisions as to whether a robot possesses a given attribute and the speed with which they make these decisions.

Anthropomorphism in trait attribution may occur in several ways. People may assume that a robot possesses attributes that are indicated by its actions; e.g., a robot that acts as if it is sensitive to its listener may be deemed sensitive in nature. But for humans, trait attributions often go far beyond observable behavior. Attractive people, for example, are thought to possess a variety of other unrelated positive traits (e.g., intelligence, friendliness). In addition, people may apply person schemata or stereotypes to people based on small bits of behavior, and use these schemata to draw assumptions about a wide range of traits, both positive and negative. (See [23] for a review).

As discussed previously, there is evidence that people will in fact attribute human personality traits to robots based on such factors as voice [15], facial features [21], and ability to take the listener's perspective [28]. Open questions include whether they attribute these properties to robots to the same extent they do for a human enacting the same behavior, and whether they also attribute robotic traits simultaneously.

In addition to looking at whether participants think that a specific characteristic is true or false of a robot vs. a human enacting the same role, we can also examine how quickly they make this decision. Research in cognitive psychology has shown that people's reaction times for binary decisions, such as whether a given adjective describes a target person, are faster when the stimulus word has been previously primed (cf. [19]). If reading a story about a robot primes anthropomorphic conceptualizations more than machine-like conceptualizations, reaction times for adjectives for human personality traits (e.g., *caring, attentive*) should be faster than reaction times for adjectives reflecting machine attributes (e.g., *artificial, software*).

We also explore the possibility of using reaction times to distinguish metaphorical vs. literal interpretations of personality adjectives. Metaphorical expressions generally take longer to process than literal expressions, particularly when they are not primed by context (see [2] for a review). Thus, if people believe that robots literally possess personality traits, their reaction times should match those of others making the same judgments with respect to a human exhibiting the same behaviors. If instead people believe that robots only metaphorically possess personality traits, their judgments may be slower than those of others making the same judgments about a human. Because the prior literature is mixed on whether metaphoric applications of adjectives will be slower than literal applications, we formulated our investigation of reaction times as a research question only.

1.3 Level 3: Abstract Judgments

In contrast to the linguistic and attribution measures previously discussed, respondents in HRI studies may also be asked to provide judgments about robots in general. For example, they might be asked whether robots experience feelings and emotion or whether robots have needs and desires. Barrett and Keil [1] showed that people's judgments about novel entities such as God or a futuristic robot varied depending on the depth of thought they give to the judgment. In their spontaneous interpretations of events, people readily anthropomorphized whereas upon deeper reflection they showed cognizance of the differences between these novel entities and human beings.

Why might survey data differ from less explicit measures of anthropomorphism such as word choices or reaction times? An obvious reason is that survey responses have less time pressure, allowing participants to think through issues such as whether robots truly experience emotion. Another reason may be people's desires to present themselves in a socially desirable way, namely, as a rational person. Self-presentational concerns have been posited as explanations for differences between explicit and implicit measures of stereotyping and bias (e.g., [4][5]). Finally, survey responses may tap people's concepts of robots in general as opposed to their concepts of a specific robot in a specific context. Within the context of a specific interaction with a particular robot, anthropomorphism provides a handy framework for understanding a robot's behaviors and predicting its future actions. In addition, a specific robot, especially upon repeated encounters, may start to feel familiar. Familiarity, in turn, has been associated with reduced processing of social information and more stereotyped responses [27].

1.4 How Robot Behavior Might Change Anthropomorphism

Robots can exhibit behaviors that would be interpreted negatively if performed by humans. For example, robots might continue to repeat a question rather than reformulating it based on the listener's questions (see the bottom excerpt in Table 1). In humans, such behavior would be considered rude and insensitive, even possibly aggressive. But can robots possess bad personalities and bad intentions? Or could such behavior be better explained by (intentional or unintentional) errors in the software driving the robot? There is some evidence that people do anthropomorphize robot behavior that is socially inappropriate or insensitive. When robots make errors in perspective-taking by providing more or les information than the listener needs, listeners are more likely to say that it is patronizing [28].

But whereas people anthropomorphize negative robot behavior, they may do so less than they do positive robot behavior. Research on human social cognition suggests that people attribute more complex human qualities to people they like [11]. Similarly, people view positive behaviors like fetching in dogs as more intentional than negative ones, such as having an accident in the house ([8], study 1). And liking for a pet fish was associated with greater anthropomorphism using both communication measures and trait attributions ([8], study 3).

1.5 Hypotheses

Based on the discussion above, we posited two hypotheses about anthropomorphism in robot perception:

Hypothesis 1: People will anthropomorphize robots more in their responses to a particular robot in a specific context (Levels 1 and 2) than in their responses to robots in the abstract (Level 3).

Hypothesis 2: People will anthropomorphize a polite robot more than an impolite robot.

2. METHOD

2.1 Overview

The study consisted of three phrases. In phase 1, participants read excerpts from a health interview and summarized them in their own words. In phase 2, they provided true or false responses as quickly as possible to a set of adjectives pertaining to the health interviewer. In phase 3, they completed an online survey in which they were asked about the characteristics of a robotic health interviewer and provided basic demographic information. The manipulation was a 2 by 2 design, between subjects. Half of the subjects thought that the interviewer in the extracts was a robot; the other half thought it was a human. Half of the subjects in each condition read extracts in which the interviewer was friendly; the other half read extracts in which he was unfriendly.

2.2 Participants

Participants consisted of 74 undergraduate students, graduate students and staff members from local universities (62% male, 68% under age 25). Of these, 62% identified themselves as white, 16% as African American, 16% as Asian, 4% as Latino, and 1% as other. All were native English speakers or fluent in English.

2.3 Materials

Vignettes. Vignettes were adapted from a previous study in which a robot interviewed participants about their health-related activities [22]. We chose seven excerpts from this study for the current experiment. In each extract, the robot asked a different student about an aspect of his/her health-related behavior, such as, "do you spend time exercising?" or "what is your favorite food?" For two of the extracts, we altered the text to create an *impolite* version and a *polite* version. An example is shown in Table 1.

Table 1. Example of excerpt from the polite interviewer and impolite interviewer conditions.

Polite Interviewer Condition Interviewer: How much do you weigh, Dan? Student: [Long pause] I would rather not answer. Interviewer: I don't mean to pry if you don't want to say. Student: I really don't want to say. Interviewer: No problem! Sorry.
Impolite Interviewer Condition Interviewer: How much do you weigh, Dan? Student: [Long pause] I would rather not answer. Interviewer: How much do you weigh, Dan? Student: I really don't want to say. Interviewer: No problem! Can I weigh you?

A page introducing the robotic or human interviewer, named Bob, was created. The page included an image of Bob (Figure 1) to highlight the interviewer condition to the participants.

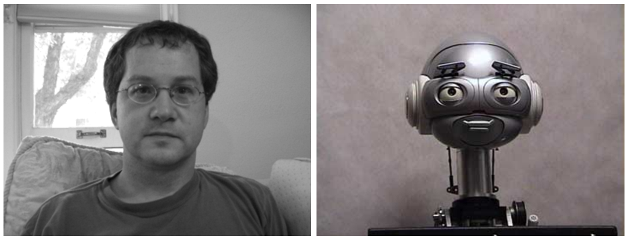

Figure 1. Pictures of the Human and Robotic Interviewers

Adjective lists. Sets of adjectives were created based on prior research ([7] [12]). Adjectives fell into four sets of ten items each, as shown in Table 2. The first set of adjectives reflects the sociability dimension. Half of these were on the positive end of the dimension (e.g., *friendly*, *polite*) and therefore true of the polite interviewer and false of the impolite interviewer; the other half were on the negative end (e.g., *rude*, *obnoxious*) and therefore false of the polite interviewer and true of the impolite interviewer. The second set of adjectives consisted of other human personality traits or states. Half of these were true for both polite and impolite human interviewers (e.g., *organized*, *curious*) and the other half were false for both polite and impolite human interviewers (e.g., *nervous*, *distractible*). The third set of adjectives consisted of robot-specific terms (e.g., *breakable*, *robotic*) that were true of the robot but false of the human interviewers. Finally, the fourth set of adjectives consisted of terms that were false for both the human and robotic interviewers (e.g., *wet*, *roasted*).

Table 2. Adjectives used in the true/false phase of the experiment

Category	Examples
Human Sociability	Friendly, polite, sensitive, caring, sociable
	Rude, obnoxious, cold, impatient, aggressive
Other Human	Organized, thorough, male, curious, persistent
	Nervous, distractible, shallow, female, disorganized
Robotic	Android, artificial, automaton, mechanical, synthetic, breakable, controllable, robotic, software, portable
False fillers	Animal, wooden, wet, smelly, tubular, ceramic, cotton, striped, roasted, bloody

Survey. An online post-task survey was created for use in Phase 3 of the study. This survey contained a manipulation check in which participants indicated whether they had read about a human vs. a robotic interviewer, and whether they had read about a polite vs. an impolite interviewer. The survey also contained 16 yes/no questions about robot's attitudes, moods and feelings (Table 5 below) and basic demographic questions (age, gender, academic major, etc.)

Software. The reaction time phase of the study was run using Empirisoft's MediaLab software (www.empirisoft.com). The software introduced the human or robotic interviewer (using the pictures in Figure 1), and then presented the vignettes to be summarized followed by the true/false adjective list. The order of presentation for the adjectives was randomized automatically by the system. Respondents used the 1 and 2 keys on the keypad to indicate true and false judgments, respectively. Timing was recorded in milliseconds.

2.4 Procedure

Participants were run individually. Upon arrival at the lab, they were seated at a computer running MediaLab, provided with an overview of the study, and then asked to sign a consent form. They then began Phase 1 of the study. They were presented with the six vignettes one at a time. After each, they had to summarize the interaction in their own words on a new page that prevented them from reading back. After completing Phase 1, they moved on to the true/false questions in Phase 2. They first practiced using the keys by making 5 true and false judgments (e.g., "grass is blue"). Then, they saw the 40 adjectives in randomly presented order. For each, they decided as quickly as possible whether the adjective was true or false for the interviewer in the extracts they had read. Upon completion of the true/false items, participants did the online survey. Afterward, they were thanked, told about the purpose of the study, paid, and dismissed.

2.5 Measures

We assessed four sets of dependent measures: language use in the vignette summaries, true vs. false decisions in the reaction time segment of the study, reaction times for these decisions, and responses to the post-task questionnaire.

Language use. Participants' summarizations of the vignettes they had read were processed using Kramer et al.'s TAWC software [9]. TAWC automatically identified instances in four major categories of interest, drawn from Pennebaker et al.'s Linguistic Inquiru and Word Count tool [17]: negative emotions, positive emotions, cognitive processes, and social processes (See Table 3). In addition, we counted first, second and third-person pronouns as well as choices of referring expressions for the interviewer (e.g., Bob, the robot, the interviewer).

Table 3. Linguistic categories used in the current study.

Word Category	Examples
Negative Emotions	Worthless, hate, tense
Positive Emotions	Joy, love, good
Cognitive Processes	Know, think, consider
Social Processes	Converse, share, friends

In addition, participants' descriptions of the interviewer's behavior were coded using the Linguistic Category Model manual [3]. First, we extracted all verbs and adjectives related to the interviewer; then we coded each of these in terms of its category in the LCM. Examples of verbs used in the study and their classifications are shown in Table 4. We created a score for the raw number of words or phrases in each category and a score for the number of unique words or phrases in each category. We also calculated the percentage of interviewer-related verbs/adjectives in each category.

Table 4. Categories used to code descriptions of the interviewer, with definitions and examples from the corpus.

Category	Definition	Examples
Descriptive Action Verb	Single action with a physically invariant feature	talk, point, laugh
Interpretive Action Verb	Single action without a physically invariant feature Emotional consequences of a single action action	urge, tempt, probe
State Verb	Cognitive or emotional state with no clear beginning or end	feel, think, want
Adjective	Characteristic or feature of a person	pushy, appreciative

True-false responses. Participants' average "true" responses for each of the four sets of adjectives (politeness, other human, mechanical, and nonhuman) were computed after inverting negative terms such as "rude". Participants received a single score for each of the four categories.

Reaction times. Participants' response times for each true/false item were recorded in milliseconds. A preliminary analysis of the data showed that average reaction time increased linearly with age (F [1, 73] =5.65; p = .02). To enable comparisons between participants, we therefore centered reaction times by subtracting the participants' overall mean time. These centered values were normally distributed and used in all statistical comparisons.

Survey responses. Participants' yes/no responses to the 16 post-task survey questions were scored for correctness using the key shown below in Table 5. Participants received a score for the total of their correct responses.

3. RESULTS

The study provides three sets of results that shed light on people's anthropomorphism of robots. We first look at the language people used in their summaries of the robotic interview excerpts, to see how it compares to that of participants summarizing the same interviews attributed to a human interviewer. Then, we examine participants' true and false decisions for adjectives pertaining to the interviewers' behavior and the speed of their reaction times. Finally, we examine their responses to the post task survey about the characteristics of robots in general.

3.1 Free Description (Level 1)

After reading each of the seven vignettes, participants summarized them in their own words. In terms of the words used to describe the vignettes, there was little difference between those who read about a robotic interviewer and those who read about a human interviewer. Across all vignettes, participants in the two conditions used nearly identical numbers of words (for the human condition, M = 248, SD = 90; for the robot condition, M = 243, SD = 102, F < 1, *ns*). We next examined proportions of words for positive emotions, negative emotions, cognitive mechanisms, and social interaction. As shown in Figure 2, there were no significant differences in percentages of words in each category across conditions (MANOVA F < 1, ns.).

As described earlier, the words used to describe the interviewers' behavior can range in linguistic abstraction from descriptive action verbs (most concrete) to adjectives (most abstract). More abstract descriptions of a robot's behavior suggest anthropomorphism. We analyzed proportions of descriptive action verbs, interpretive action verbs, state verbs and adjectives in participants' descriptions of the interviewer's behavior. There was no main effect of interviewer condition—participants who summarized vignettes about a robot interviewer used as many abstract verbs and adjectives as participants who summarized vignettes about a human interviewer (Multivariate F < 1). There was a significant effect of interviewer politeness. As shown in Figure 3, more interpretive action verbs were used for less polite interviewers, both robotic and human (F [1, 64] = 4.52, p < .07, η^2=.06[1]). There was no interaction between interviewer condition and interviewer personality. Clearly, people describe a robotic interviewer's behavior using the same social language as they use to describe a human interviewer, consistent with Hypothesis 1. However, there was no evidence from the linguistic data that people anthropomorphize positive behaviors more than negative behaviors as predicted in Hypothesis 2; instead, they use more interpretive terms for negative behaviors.

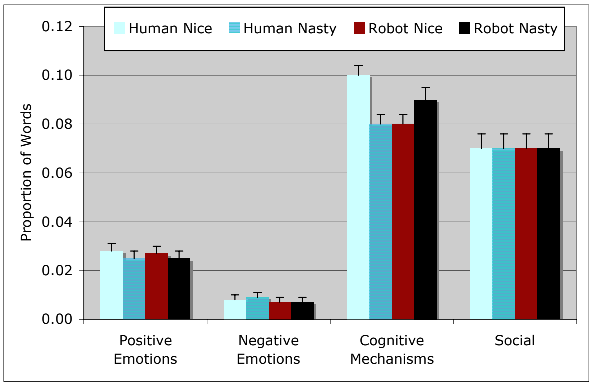

Figure 2. Proportion of words in four categories by interviewer condition (Level 1)

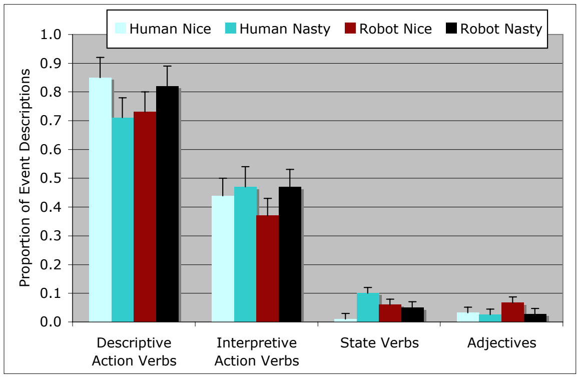

Figure 3. Proportion of free descriptions containing direct action verbs, interpretive action verbs, state verbs and adjectives by interviewer condition (Level 1)

3.2 Attributions and Reaction Times (Level 2)

In the reaction time component of the study, participants responded "true" or "false" as quickly as possible. We analyzed both "true" responses and reaction times using repeated measures ANOVAs in which interviewer condition (human vs. robot) and interviewer politeness (polite vs. impolite) were between-subjects variables and adjective category (friendliness, other human, robotic and nonhuman) was a within-subjects variable.

As expected, there was a significant main effect for adjective category (F [3, 192] = 90.05, p < .001, η^2=.53). People were much more likely to respond "true" to the friendliness and other human adjectives than to the nonhuman adjectives (Figure 4). There was also a significant interviewer condition by adjective category interaction (F [3, 192] = 9.58, p < .001, η^2=.06). Post-hoc contrasts revealed that the effect of interviewer condition was significant only for the robotic adjectives (F [1, 64] = 40.96, p < .001, η^2=.10); for all other adjective categories, responses for robot and human interviewers did not differ. Thus, these results are consistent with Hypothesis 2: people attributed human personality traits to the robot described in the vignette.

[1] Eta squared (η^2) is a measure of effect size appropriate for multivariate designs [16]. Eta squared indicates the proportion of variance in the dependent measure accounted for by an effect. Larger values indicate larger impact.

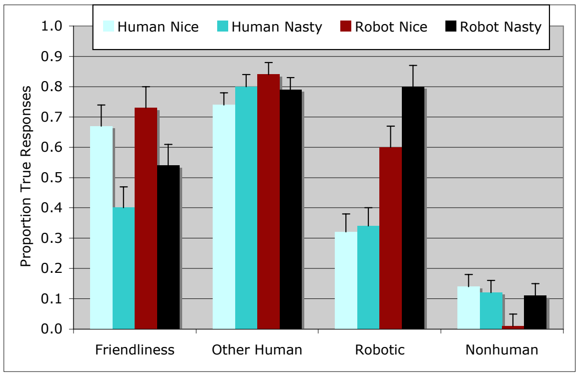

Figure 4. Proportion of true responses by condition and adjective category (Level 2).

There was also a significant interviewer politeness by adjective category interaction (*F* [3, 192] = 5.32, *p* < .005, η^2=.03), shown in Figure 4. People were more likely to attribute positive traits to the polite interviewer and negative traits to the impolite interviewer, regardless of whether the interviewer was a human or robot (*F* [1, 64] = 4.49, *p* <.05, η^2=.03). This suggests that participants are fine-tuning their anthropomorphic judgments based on their overall evaluation of the robot, but contrary to Hypothesis 2, there was no sign that people anthropomorphized the polite robot more than the impolite robot.

There was a borderline significant tendency for people to attribute the robotic adjectives to impolite interviewers (*F* [1, 64] = 3.62, *p* = .06, η^2=.01), perhaps because the robotic terms can be used metaphorically to suggest negative traits in humans. There was no three-way interaction between interviewer condition, interviewer politeness, and adjective category.

We next examined participants' reaction times for each adjective set using a 2 (human vs. robot interviewer) by 2 (polite vs. impolite interviewer) by 4 (adjective category) repeated measure ANOVA. There was a significant effect of adjective category (*F* [3, 192] = 3.32, *p* < .05, η^2=.05) but no other main effects or interactions. Participants in both the human and robotic interviewer condition responded faster to the friendliness and nonhuman adjectives than to the other two sets of adjectives.

3.3 Abstract Judgments (Level 3)

Participants responded "yes" or "no" to 16 questions about the characteristics of a generic robot interviewer, as opposed to the specific robot interviewer rated in the previous section. These questions were intended to elicit careful thought about the abstract characteristics of robots in general.

We analyzed the number of questions answered correctly by each subject, in a 2 (human vs. robot interviewer) by 2 (polite vs. impolite interviewer) ANOVA. In general, participants showed much less anthropomorphism in their abstract ratings than they did on the communication and attribution measures, consistent with Hypothesis 1.

Contrary to our expectations, there was a trend for participants in the robot interviewer condition to score lower on this test (*M* = 13.31, *SEM*=.39) than participants in the human interviewer condition (*M* = 14.21, *SEM* =.37; *F* [1, 68] = 2.84, *p* = .10, η^2=.04). To better understand this trend, we examined means for each question as a function of interviewer condition (see Table 5). Participants in the robot condition were (nonsignificantly) more likely to erroneously agree that robots had moods and feelings, liked people after getting to know them, and experienced frustration.

Table 5. Proportion of participants in each interviewer condition agreeing that a robot possesses abstract properties (Level 3).

	Human	Robot
Does a robot follow a script? (Yes)	.97	1.00
Is it likely a robot would ever need repair? (Yes)	.98	.87
Do robots ever wear out? (Yes)	.88	.85
If a robot acts happy today is it likely to act happy tomorrow? (Yes)	.91	.74
Can a robot make mistakes? (Yes)	.95	.90
Is a robot impartial? (Yes)	.63	.70
Could a robot want to change the subject? (No)	.67	.65
After getting to know someone would the robot like that person? (No)	.85	.74
Can a robot have moods? (No)	.86	.79
Would a robot ever make a mistake on purpose? (No)	.82	.88
Does a robot experience frustration? (No)	.92	.81
Would a robot feel bad if it made a mistake? (No)	.91	.85
Does a robot ever get tired? (No)	.92	.91
Does a robot care about its appearance? (No)	.97	.88
Can a robot imagine things it has not learned? (No)	.96	.89
Does a robot have feelings? (No)	.97	.89

4. DISCUSSION

The results clearly demonstrate a disjuncture between anthropomorphism in people's spontaneous reactions to robots in social context and anthropomorphism in their more carefully considered conceptions of robots in the abstract. In participants' free form summaries of health interviews, they used words to express positive and negative emotions, cognitive mechanisms, and social interaction equally much for robotic and human interviewers. Participants also used abstract interpretive verbs to describe robot behaviors about as often as they did human behaviors.

In their spontaneous trait judgments, people attributed human characteristics as often, and as rapidly, to a robot interviewer as to a human interviewer. Both the robot and human were perceived as organized and curious but not nervous or distractible, in keeping with their behaviors in the vignettes. Furthermore, people fine-tuned their assessments of a robot's personality in much the same way as they do for people. Nice people and robots were seen as friendly and polite, whereas mean people and robots were seen as rude and obnoxious. Combined with the communication data, this is strong evidence that people conceptualized the robotic interviewer anthropomorphically.

At the same time, however, people appear to hold a parallel conception of the robotic interviewer as a mechanical entity. They attributed properties such as *breakable*, *controllable*, *synthetic* and *mechanical* frequently to the robotic interviewer but rarely to the human interviewer. For robots (and people), being impolite resulted in more "true" judgments for robotic terms. Participants made judgments about the truth of human friendliness adjectives and robotic adjectives equally rapidly, suggesting that people may hold both humanistic and mechanistic conceptions of the robot simultaneously.

While people's spontaneous decisions about a robot's attributes showed considerable anthropomorphism, their more carefully reasoned responses in the post-task questionnaire revealed a much more mechanical view. Here, the vast majority of participants in both conditions denied that a robot could have moods, experience frustration, or possess feelings, much as Barrett & Keil's [1] participants expressed a less anthropomorphic view of God or a futuristic robot in their post-task questionnaire.

An unexpected finding was that although participants in the robot condition showed a more mechanical view in their survey responses, they still anthropomorphized a bit. More participants in that condition agreed that a robot might act happy one day but not the next, that a robot would like someone after getting to know him or her, that a robot had moods, and that a robot could experience frustration. It is possible that people's attitudes about robots were shaped by their reading of the vignettes (in which, e.g., a robot seems to be expressing frustration), or by their own anthropomorphic word use when describing these vignettes. Such results would be consistent with other research showing that people's language affects their cognitive processes (e.g., [18], [6], [24]). If this interpretation is correct, we might expect people's abstract conceptualizations of robots to become more and more anthropomorphic as robots penetrate daily life and daily conversation.

A question remains as to whether participants' judgments in the reaction time study reflect beliefs that robots literally possess feelings, attitudes and personality traits or whether they instead are based on metaphoric extension. This question cannot be determined by reaction time data alone, as studies have shown that evaluations of metaphoric statements like "my surgeon is a butcher" can be as rapid as that of literal statements like "John is a butcher", especially when the prior context supports the metaphorical interpretation (e.g., [2]).

One limitation of the current study is that people summarized vignettes describing interactions between the robot and another person; they did not interact directly with the robot. In future work, we intend to examine the extent to which people's responses to a robot are driven by the anthropomorphic view evidenced in their communication and trait attributions vs. the mechanistic view evidenced in their survey responses. A second limitation is that people only read about the interactions, they could not watch them. In a follow up study, we plan to replicate this experiment using short video clips of a human vs. a robot interviewer asking students about their health.

The results have implications for research methods in human-robot interaction. Specifically, our findings suggest that people's beliefs about robots as expressed in surveys removed from any particular interaction with a robot may not match the beliefs that actually guide their behavior when interacting with a robot. To understand these guiding beliefs, researchers will need to use linguistic measures, reaction times, or other measures of psychological processes.

5. CONCLUSION

In this study we analyzed anthropomorphism in people's reactions to a robot in social context vs. their thoughts about robots in the abstract. As predicted, participants were significantly more anthropomorphic in verbal descriptions of a specific robot's behavior and in their judgments of that robot's personality characteristics than they were in their judgments of robots in general. Contrary to predictions, however, people were no more or less anthropomorphic with robots exhibiting negative behaviors than with those exhibiting positive behaviors. The results have implications for the design of HRI studies that seek to understand people's conceptualizations of robots.

6. ACKNOWLEDGMENTS

This research was funded by NSF grants CNS-0551554 and HSD IIS-0624275. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation. We thank Hao-Chuan Wang for helping with the TAWC analyses and Tom Perkins for posing as our human interviewer.

7. REFERENCES

[1] Barrett, J. L. & Keil, F. C. (1996). Conceptualizing a nonnatural entity: Anthropomorphism in God concepts. *Cognitive Psychology, 31,* 219-247.

[2] Cacciari, C., & Glucksberg, S. (1994). Understanding figurative language. In M. A. Gernsbacher (Ed.), *Handbook of psycholinguistics* (pp. 447-477). New York: Academic Press.

[3] Coenen, L. H. M., HIdebouw, L., & Semin, G. (2006). *The Linguistic Category Model (LCM) Manual* (parts 1 and 2). June 2006 Version.

[4] Dovidio, J. F., Kawakami, K., & Beach, K. R. (2001). Implicit and explicit attitudes: Examination of the relationship between measures of intergroup bias. In R. Brown, & S. L. Gaertner (Eds.), *Blackwell handbook on social psychology (Vol. 4, pp. 175–197). Intergroup relations*. Oxford: Blackwell.

[5] Greenwald, A. G., Banaji, M. R., Rudman, L. A., Farnham, S. D., Nosek, B. A., & Mellott, D. S. (2002). A unified theory of implicit attitudes, stereotypes, self-esteem, and self-concept. *Psychological Review, 109,* 3–25.

[6] Higgins, E. T., & Rholes, (1978). "Saying is believing": Effects of message modification on memory and liking of the person described. *Journal of Experimental Social Psychology, 14,* 363-378.

[7] Kiesler, S., & Goetz, J. (2002). *Machine trait scales for evaluating mechanistic mental models of robots and computer-based machines.* Unpublished manuscript, Carnegie Mellon University. Downloadable at http://anthropomorphism.org/pdf/Machine_scale.pdf

[8] Kiesler, S., Lee, S-L, & Kramer, A. D. I. (2006). Relationship effects in psychological explanations of nonhuman behavior. *Anthrozoös, 19,* 335-352.

[9] Kramer, A. D. I., Fussell, S. R., & Setlock, L. D. (2004). Text analysis as a tool for analyzing conversation in online support groups. *CHI 2004 Late Breaking Results* (pp. 1485-1488). NY: ACM Press.

[10] Lee, S., Kiesler, S., Lau, I.Y. & Chiu, C-Y. (2005). Human mental models of humanoid robots. *Proceedings of the 2005 IEEE International Conference on Robotics and Automation (ICRA '05)*. Barcelona, April 18-22., 2767-2772.

[11] Leyens, J., Paladino, P. M., Rodriguez-Torres, R., Vaes, J., Demoulin, S., Rodriguez-Perez, A., & Gaunt, R. (2000). The emotional side of prejudice: The attribution of secondary emotions to ingroups and outgroups. *Personality and Social Psychology Review, 4,* 186-197.

[12] Loughnan, S., & Haslan, N. (2007). Animals and androids: Implicit associations between social categories and nonhumans. *Psychological Science, 18,* 116-121.

[13] Morkes, J., Kernal, H. K., & Nass, C. (1999). Effect of humor in task-oriented human-computer interaction and computer-mediated communication: A direct test of SRCT theory. *Human-Computer interaction, 14,* 395-435.

[14] Nass, C. & Brave, S. (2005). *Wired for speech: How voice activates and advances the human-computer relationship*. Cambridge, MA: MIT Press.

[15] Nass, C. & Lee, K. M. (2001). Does computer-synthesized speech manifest personality? Experimental tests of recognition, similarity-attraction, and consistency-attraction. *Journal of Experimental Psychology: Applied*, *7*, 171-181.

[16] Olejnik, S. & ALgina, J. (2003). Generalized eta and omega squared statistics: Measures of effect size for some common research designs. *Psychological Methods, 8,* 434-447.

[17] Pennebaker, J.W., Francis, M.E., & Booth, R. J. (2001). *Linguistic Inquiry and Word Count: LIWC (2nd ed.).* Mahwah, NJ: Lawrence Erlbaum Associates.

[18] Pennebaker, J. W., Mehl, M. R., & Niederhoffer, K. (2003). Psychological aspects of natural language use: Our words, our selves. *Annual Review of Psychology, 54,* 547-577.

[19] Posner, M. I. (1978). *Chronometric explorations of mind.* Hillsdale, NJ: Erlbaum.

[20] Powers, A., Kramer, A. D. I., Lim, S., Kuo, J., Lee, S-L., & Kiesler, S. (2005). Eliciting information from people with a gendered humanoid robot. *Proceedings of the 14th IEEE International Workshop on Robot and Human Interactive Communication (*ROMAN 2005).

[21] Powers, A., & Kiesler, S. (2006). The advisor robot: Tracing people's mental model from a robot's physical attributes. *Conference on Human-Robot Interaction 2006*. Salt Lake City, March 1-3, 218-225.

[22] Powers, A., Kiesler, S., Fussell, S., & Torrey, C. (2007). Comparing a computer agent with a humanoid robot. *Proceedings of HRI07*, pp. 145-152.

[23] Quinn, K. A., Macrae, C. N., & Bodenhausen, G. V. (2003). Stereotyping and impression formation: How categorical thinking shapes person perception. In M. A. Hogg & J. Cooper (Eds.), *Sage handbook of social psychology* (pp. 87-109). Thousand Oaks, CA: Sage Publications.

[24] Schooler, J. W., & Engsler-Schooler, T. Y. (1990). Verbal overshadowing of visual memories: Some things are better left unsaid. *Cognitive Psychology, 22*, 36-71.

[25] Semin, G. R., & Fiedler, K. (1988). The cognitive functions of linguistic categories in describing persons: Social cognition and language. *Journal of Personality and Social Psychology, 54,* 558-568.

[26] Semin, G. R., & Fiedler, K. (1991). The linguistic category model, its bases, applications and range. In W. Stroebe & M. Hewstone (Eds.), *European review of social psychology* (Vol. 2, pp.1-30). Chichester, England: Wiley.

[27] Smith, E. R.,' Miller , D. A., Maitner, A. T., Crump, S. A., Garcia-Marques, T., Mackie, D. M. (2006). Familiarity can increase stereotyping**.** *Journal Of Experimental Social Psychology***,** *42,* 471-478.

[28] Torrey, C. Powers, A., Marge, M., Fussell, S. R., & Kiesler, S. (2006). Effects of adaptive robot dialogue on information exchange and social relation. *Proceedings of the Conference on Human-Robot Interaction 2006*, pp. 126-133.

[29] Wigboldus, D. H. J., Dijksterhuis, A., & Van Knippenberg, A. (2003). When stereotypes get in the way: Stereotypes obstruct stereotype-inconsistent trait inferences. *Journal of Personality and Social Psychology, 84*, 470-484.

[30] Yee, N., Bailenson, J.N., & Rickertsen, K. (2007). A meta-analysis of the impact of the inclusion and realism of human-like faces on user experiences in interfaces. In *Proceedings of the Conference on Human Computer Systems* CHI'07. pp. 1-10, NY: ACM Press.