Mental Models and Cooperation with Robotic Assistants

Sara Kiesler
Human Computer Interaction Institute
Carnegie Mellon University
Pittsburgh, PA 15213
412-268-2848
kiesler@cs.cmu.edu

Jennifer Goetz
Human Computer Interaction Institute
Carnegie Mellon University
Pittsburgh, PA 15213
412 268-5919
jenga@andrew.cmu.edu

ABSTRACT
In the future, interactive robots will perform many helpful tasks. In 5 studies, we developed techniques for measuring the richness and content of people's mental models of a robot. Using these techniques, we examined how a robot's appearance and dialogue affected people's responses. Participants had a comparatively rich mechanistic perception of the robot, and perceived it to have some human traits, but not complex human attachment, foibles, or creativity. In study 5, participants who interacted with an extraverted, playful robot versus a more serious, caring robot, developed a richer, more positive mental model of the playful robot but cooperated less with it. Our findings imply different designs for robotic assistants that meet social and practical goals.

Keywords
Experiment, human-robot interaction, mental model, anthropomorphism, cooperation, compliance, robotic assistants.

INTRODUCTION
Advances in computer technology, artificial intelligence, speech simulation and understanding, and remote controls have led to breakthroughs in autonomous mobile robots. Robots can identify and track a user's position, respond to spoken questions, display text information, and travel on command while avoiding obstacles. In the future, robots will entertain people and assist in a range of tasks that are unpleasant, unsafe, taxing, confusing, low paid, or boring to human assistants. However, these robots must meet social as well as instrumental goals. They must create a comfortable experience for people, gain their cooperation, encourage healthy rather than overly dependent behaviors, and provide appropriate feedback to remote operators and others involved in the system. Although researchers are gaining practical experience with mobile, autonomous robots in settings such as museums [22], we lack a principled understanding of how to design robots that will accomplish these social goals. Toward that end, we have begun to conduct systematic research on how people perceive and interact with robotic assistants.

Theoretical Framework
Mental models are a term for the conceptual frameworks that support people's predictions and coordination in a dynamic world. People form anthropomorphic mental models of higher animals, deities, nature, and animated objects and machines (e.g., [12, 15, 18]). Anthropomorphic mental models seem to be universal and intuitive, a kind of built-in cognitive default even among scientists who argue these attributions are inappropriate [4]. In 1946, Hebb [11] proposed that anthropomorphic descriptions help scientists cognitively organize and predict the behavior of primates, e.g., [10]. In the same vein, anthropomorphic mental models could help people understand and frame their responses to autonomous robotic assistants.

Mental models develop and change with experience [8] but it is not known how mental models of robotic assistants (or other technologies) become more or less anthropomorphic. Perhaps, if these robots have salient humanlike attributes such as speech and purposeful movement, people's mental models of robotic assistants will become more anthropomorphic as they interact with them.

Researchers working on believable agents and emotional robots [1, 3, 17, 20] have drawn on the traditions of animation, drama, and screenwriting to create engaging computer agents that seem lifelike to viewers. However, the cognitive processes involved in anthropomorphic mental models have not been well understood. How do people form a coherent mental model of an object they know to be a machine but that nonetheless exhibits what looks like intention, emotionality, or otherwise humanlike behavior? Theories of instance-based cognitive processing [13] and exemplar-based processing [16] suggest people can integrate ostensibly incompatible images and categories into a consistent, anthropomorphic mental model. For example, a life-like robot that tells a joke could activate in the viewer's memory exemplars of the nonsocial category, *machines*, and of the social category, *humorous people*. Combining these exemplars could lead to the experience of an integrated concept, such as *cheerful robot*. If so, humanlikeness in a machine, either by virtue of its appearance or its behavior, can lead to a mental model that does not deny the technology in the machine, but that also incorporates anthropomorphic features into it.

Overall research strategy

Our goal is to understand the anthropomorphic process. To start research towards that goal we carried out a series of initial studies of the influence of a robot's appearance and dialogue on how people think about the robot and act towards it. We conducted 3 studies with an interactive toy robot; we conducted our later work with "Pearl," an autonomous interactive robot.

An intermediate goal of our work has been to develop measures of the richness and content of people's mental models of robots on anthropomorphic as well as mechanistic dimensions. We believe this work will aid assessment of people's trust in interactive robots, and the degree to which they have appropriate conceptions of them.

Richness of the mental model

Prior research shows that people have a sparse, simple mental model of members of out-groups as compared with their own groups [7] and with those with whom they have little direct experience [9]. As people interact with others, their mental models become richer, as shown by their giving more trait ratings [7] or by their responding quickly or confidently when making these ratings [9, 21].

We have built on this literature to create measurements of robot mental model richness and confidence. In studies 1 and 2, we adapted rating scales of human social evaluation (5 items) and intellectual evaluation (4 items) from [23] and used rating scales of "humanlikeness," each of which have been used previously in research on computer agents (e.g., [19]). In studies 1, 2, and 3 we used the number of trait ratings to measure richness of the mental model; in study 5 we used extremity of ratings and response times.

Content of the mental model

In our later studies, we wished to explore the nature of people's mental models, rich or sparse. We therefore expanded the social evaluation scale to explore personality dimensions of anthropomorphism in greater detail, so we adapted the Big Five Inventory from [14] in addition to the intellectual evaluation and humanlikeness items. In study 3, we used rating scales of mechanistic mental models, created for that study.

Cooperation – behavioral measures

In the future, robots that perform helpful tasks will need not only to engender trust, but also to obtain the cooperation of their users. For example, if the robot is assisting in the care of elderly people, a design goal could be inducing compliance with medication reminders, exercise suggestions, or meal delivery times. Even now, autonomous mobile robots carrying supplies in hospitals request people to relinquish their place on elevators. Hence, we are developing measures of compliance and cooperation. In study 1, we asked people to complete long questionnaires and measured their willingness to do so [5]. In study 4 (a user test with the Pearl robot) and in study 5 (a lab experiment with the Pearl robot), we measured how long people would be willing to perform a physical exercise for the robot.

STUDIES WITH INTERACTIVE TOY ROBOTS

We constructed two plastic toy robots (Robotix Vox Centurion™) for this work. One robot was constructed as a man, standing about 3 feet high, and one robot was constructed as a vehicle, about 3 feet long. We used a quasi-Wizard of Oz procedure to implement interaction with participants. We placed small speakers on both robots, with a wireless connection to a remote laptop, and used WillowTALK 4.0 software for the robot's speech generation, with the "Paul" voice. The robot gave a standardized script, with little branching. The robot gave the appearance of recognizing simple speech, as the experimenter controlled the robot's script and movement through the laptop.

Study 1: Field Experiment in a Science Museum

We placed the toy robots, one by one, in an open hallway of a large science museum on two weekend afternoons. Each robot took a turn alone at two-hour intervals, standing near a table on which was piled some questionnaires. Many other interesting and distracting exhibits surrounded this spot.

Figure 1. Toy Robot-Vehicle.

Dialogue

In both conditions (humanlike or vehicle-like), the robot that was visible during its designated time period turned toward any visitor who approached, and said "hi" to engage the visitor in conversation. If the visitor replied, the robot answered, "My name is Bob. What is your name?" The robot then said, "It's nice to meet you. Would you like to do me a favor?" If the visitor responded positively or asked about the favor, the robot said, "I'm trying to learn about humans, and I have these questionnaires for people to fill out."

Measures

Our dependent variable was visitors' willingness to complete the long questionnaire.

Results

Approximately 50 people passed close enough to each robot to hear it, but compliance was comparatively low. Only 12 people completed the entire questionnaire for the robot-man. Only two people completed the entire questionnaire for the robot-vehicle. Indeed the robot-vehicle attracted few visitors' attention at all. One visitor, on hearing the robot-vehicle say "hi", walked directly past it and looked behind our barrier to speak to the robot-man. These responses suggest a robot with a humanlike appearance is more likely to engage people initially than a robot that is not humanlike. The head of a humanlike robot, with mouth and eyes, offers a focal point for people's visual, verbal, and auditory attention, and prepares the ground for future interaction.

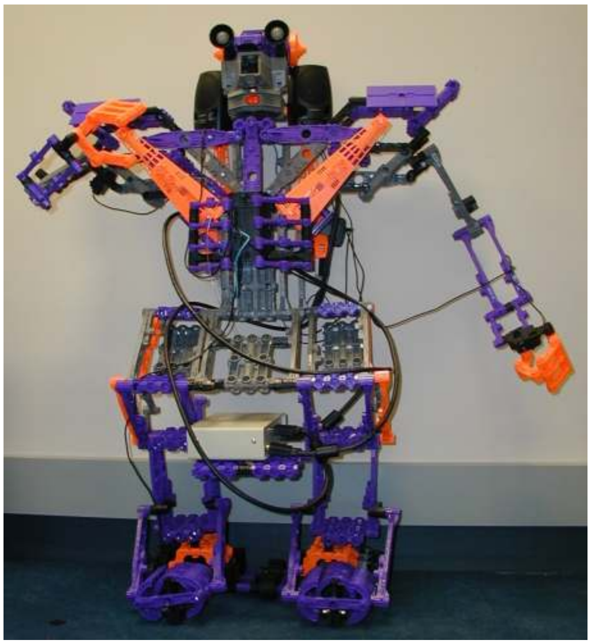

Figure 2. Toy Robot-Man.

Study 2: Toy Robot-Vehicle vs. Toy Robot-Man

For improved experimental control, we moved into the laboratory. In the first experiment, we asked college student volunteers to converse with the toy robot-vehicle (n = 10) or robot-man (n = 11).

We hypothesized that participants who interacted with the robot-man would have a richer, more anthropomorphic mental model of it than would those who interacted with the robot-vehicle. As noted above, in this study we used scales for social and intellectual evaluation and ratings of human likeness to assess the content of the mental model, and the number of trait ratings to assess the richness of the mental model.

An experimenter introduced the participant to the robot as follows: "Now I will turn on the robot and you can talk to it. It will ask you questions and you can ask it questions. If you don't understand something, please direct your question to the robot. Okay? It starts talking once I turn it on." A second experimenter, hidden in an adjacent room, listened to the participant and controlled the robot's simulated speech script.

Dialogue

As in the field experiment, the robot engaged participants in a conversation, but in this case we created a more engaging script; the robot told a joke, showed feelings, and briefly "interviewed" the participant about his or her sociability. For example:

R: *Hi, how are you today?*
P: *I'm fine, thanks. How are you?*
R: *I'm doing alright. I really appreciate you coming here to visit me. It gets so lonely here in the lab.*
P: *I'm sure it does. I thought it would be interesting to talk to you.*
R: *Do you want to hear a joke?*
P: *Yeah, sure.*
R: *Why did the turkey cross the road?*
P: *I have no idea.*
R: *Because he wasn't chicken. Ha-ha.*
P: *Hahaha.*
R: *Wasn't that funny?*
P: *That was very funny. I'm not really a joke person, but your jokes are great.*
R: *I am trying to learn about how humans interact. Do you mind if I ask you some questions?*
P: *Not at all.*
R*: Okay, we'll use a scale of 1 to 5 where 1 is very uncomfortable and 5 is very comfortable. How do you feel introducing yourself to new people?*
P: *Depends on the people…three.*
(dialogue continues)

In enacting this script, the robot "answered" simple questions, if asked, but generally avoided deviating from the script.

Measures

We used the social evaluation and intellectual evaluation scales, and measures of humanlikeness (described above) and additional items we created for this study to assess reactions to the robot (e.g., humorous, sympathetic, casual, attractive). Our dependent variables were mental model richness, as measured by the number of trait ratings (vs. "does not apply") on the scales, and mental model content, as measured by the ratings themselves.

Results

All of the participants were asked to attend to the robot, and given this attention, the robot's seeming ability to carry on a dialogue and express feelings dominated participants' responses, such as asking the robot questions about its attributes ("Do you know any swear words?" asked of the robot-man), and drawing up close to the robot to converse with it.

The dialogue seemed to affect participants' mental models more than the appearance of the robot. The vehicle-like or humanlike appearance of the robot did not affect participants' measured mental models differently.

The richness of participants' mental models were better developed on dimensions the participants could derive from their speech interaction with the robot. Participants made more trait ratings when rating obvious and simple traits of the robot (e.g., attractive, cheerful, humorous, casual, and sociable) than when rating more complex, less observable traits that would need to be inferred (e.g., responsible, sensible, and sympathetic). This difference, 87% vs. 63% of the items offered for rating, was highly significant ($F[1, 19] = 12.3$, $p < .01$; see Figure 3) and consistent with the literature on out-groups and sparse mental models [18].

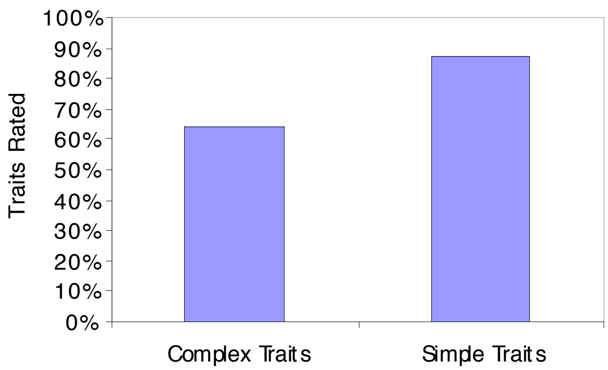

Figure 3. Propensity to rate robot on human traits.

Study 3: Adding "Hardware" to a Robot

The first experiment suggested that dialogue may dominate differences in appearance in creating rich versus sparse dimensions of a mental model. In study 3, we manipulated the robot's appearance differently to examine if appearance would change the degree of anthropomorphism in people's mental models. In one condition, we attached additional hardware—an external modem box with some cables—to the toy robot-man and in the other condition we presented the robot-man without the additional hardware. The purpose of this manipulation was to test whether the addition of hardware would alter the anthropomorphism of the mental model.

Twenty students and staff participants from Carnegie Mellon University played a simplified 5-item desert survival decision-making game [23] with the toy robot-man.

Dialogue

The robot's initial dialogue was designed to be friendly and cooperative, and more task-oriented than in study 2.

R: *I like it here. People are nice to me.*
P: *That's good.*
R: *What is your major?*
P: *Computer science.*
R: *Computer science is interesting.*
P: *Yeah, I like it.*
R: *Hey, do you want to hear a joke?*
P: *Sure.*
R: *Why did the turkey cross the road?*
P: *I don't know. Why?*
R: *Because he was not chicken. Ha-ha. Wasn't that funny?*

After the initial interaction, the experimenter introduced the task. Participants were asked to tell the robot their choices and listen to the robot give its choices. As in Experiment 1, a second remote experimenter directed the robot's dialogue. An example of the robot's script is below.

R: *Yes, I am listening. Why don't you tell me how you ranked all of the items, and then we can discuss it.*
P: *Okay, well I rated oxygen first, water second, and the food third. Then I put the map and the transmitter. How did you rank them?*
R: *I pretty much agreed with you, although I thought that the water was less important, because I assumed we would be rescued in a day or two.*

The robot script used an algorithm of always disagreeing with the participants' 2nd and 5th items, no matter what they were. The script was programmed for all possible answers. At the end of the interaction, the robot switched to agree with half of the participants and "agreed to disagree" with the other half of the participants, but this manipulation made no difference in the results and is discussed no further.

Measures

After completing the exercise with the robot, participants rated the robot and themselves. The mental model measures included the 5 scales (44 items) of the Big Five Inventory, used extensively in personality, mental health, and social psychology research. The scale items measure extraversion (e.g., talkative, enthusiastic), agreeableness (e.g., polite, helpful), conscientiousness (e.g., reliable, organized), neuroticism (e.g., moody, tense), and openness to experience (e.g., creative, artistic). We also included the intelligence evaluation scale used in the previous experiment and two humanlikeness items (looks human; acts human).

We added new measures of mechanistic mental models for this study using ratings of the following items: complex, obsolete, intuitive, works quickly, usable, durable, powerful, reliable, accurate. A factor analysis of these ratings revealed three factors (using eigenvalues over 1 to select the number of factors) accounting for 67% of the variance in responses. Items were used for a scale if their factor loading was 50% or better on that factor. The items complex, (not) obsolete, quick, intuitive, and usable loaded on the first scale, "advanced." The items reliable and accurate loaded on the second scale, "reliability." The items durable and powerful loaded on the third scale, "power."

Results

First we evaluated the richness of participants' mental models. Participants rated 100% of the mechanistic traits indicating they had a rich mental model of the robot as a machine. In contrast, as in study 2, there were sharp differences in participants' likelihood of rating anthropomorphic traits ($F[5, 190] = 13.5$, $p < .0001$). Participants were most likely to rate the robot on items measuring extraversion (rather than checking "does not apply")—93% of the time. They were least likely to rate the neuroticism and openness to experience items—68%. We think the participants created a richer, more confident mental model on extraversion of the robot because of the outward evidence of this trait in its dialogue.

Turning now to the content of the mental models, we examined the results for the items that participants rated. The appearance manipulation influenced participants' mechanistic model. The hardware manipulation caused participants to have a less positive perception of the robot's reliability (3.5 with no added hardware vs. 3. 1 with hardware; $F [1, 36]) = 3.7$, $p = .06$), a more positive perception of its power (2.1 with no hardware vs. 2.9 with hardware; $F [1, 23] = 4.7$, $p < .05$). The manipulation did not change perceptions of how advanced the robot was (1.8 in both conditions). At the same time, participants in the hardware condition had a slightly less positive perception of the robot on the anthropomorphic (Big Five) items.

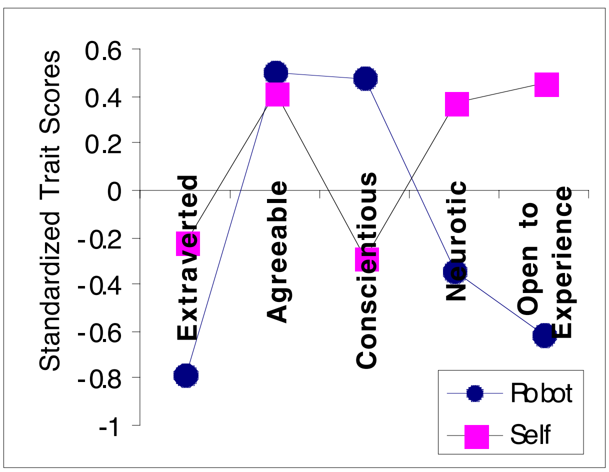

Figure 4. Ratings of robot and self (standardized for comparison).

Figure 4 shows how participants rated themselves on the Big Five personality traits as compared with how they rated the robot. We standardized these ratings so they can be compared. As shown in the figure, participants rated the robot in a different pattern than they rated themselves. The participants saw the robot as comparatively less extraverted, neurotic, and open to experience, as agreeable, and more conscientious (Trait x Robot/Self interaction $F [4, 384] = 16.6$, $p < .001$). These results suggest the limits of the anthropomorphic mental model created with this interactive toy robot. That is, the robot was seen as having some of the overtly conventional traits of humans, but not complex human attachment, foibles, or creativity.

STUDIES WITH A ROBOTIC ASSISTANT

To extent our toy robot results, we observed people's interactions with an autonomous mobile robot, a working prototype of a robotic assistant named "Pearl" (see Figure 5; http://www-2.cs.cmu.edu/~nursebot/).

The robot used in these studies was developed by a joint project in robotics, human-computer interaction, and artificial intelligence whose goal is the development of robotic assistant technology. Currently, the robot can navigate safely through indoor environments, seek out people, approach them, initiate an interaction, and respond to people's inquiries or provide reminders within a limited domain of expertise. The robot uses the Sphinx speech recognition system developed at Carnegie Mellon University, available as open source, and is capable of speaker-independent recognition using a microphone on the robot. It also uses the Festival speech synthesis system, developed by researchers at the University of Edinburgh, to emit speech. The robot also has a simple graphical interface using a touch-sensitive LCD screen. The current dialog manager handles a limited range of domains such as finding TV programs, reading the newspaper, exercising, medication taking, and obtaining weather info. The dialog manager models a person's ability to understand spoken language and respond clearly, and adds clarification questions as needed.

Figure 5. The robotic assistant.

Study 4: A User Study

Before conducting a full experiment with the robotic assistant, we conducted a user study in which the robot interacted with 3 elderly people, ages 68 to 84. Because ultimately the robot will be designed for an elderly clientele, we wished to gain an initial understanding of the interaction challenges to be faced in designing the robot to gain this group's confidence and cooperation.

In this test, the robot read a newspaper article and led each person in mild physical exercises. We observed and videotaped the interactions, noted problems and critical incidents, and interviewed the users afterwards.

The three elderly users spoke with the robot and had no greater difficulty interacting with the machine in a social context than our previous study participants did. One usability problem was that the elderly users had difficulty hearing the robot's synthetic speech [see also 2], necessitating a much higher volume. We also found some noncompliance with the robot's requests, even in this public, videotaped setting. When requested to stand up and sit down 5 times, as part of the exercise routine, the 84 year old stood up 3 times but told the robot he had completed 5 exercises. This user test confirmed our hunch that cooperation with robotic assistants cannot be taken for granted. We therefore developed a new measure of cooperation so that we could explore, experimentally, the genesis of robot mental models and cooperation within the same study.

Study 5: Playful vs. Serious Robot

In previous studies, we manipulated the appearance of a robot to examine the impact on people's mental models of the robot. In this study, we manipulated the personality of the robot through its dialogue. To extend our work to people's behavior, we examined whether cooperation with a robot would follow from people's mental model of the robot. We intended to create two comparatively rich but different robot mental models, so that we could examine the effect of mental model content on cooperation. We therefore created a serious, concerned robot personality and an extraverted, playful robot personality. We hypothesized that the mental model of a serious, concerned robot would be rich and positive on traits related to conscientiousness and agreeableness, whereas the mental model of an extraverted, playful robot would be rich and positive on traits related to extraversion and possibly, openness to experience. Since people tend to like extraverted, open others, we predicted that participants would enjoy interacting with the playful robot and cooperate best with its requests.[1]

Dialogue

Forty participants (average age = 22) interacted with the robot, shown in Figure 5. We manipulated the robot's dialogue to create the two personality types. In the serious condition, the robot repeatedly expressed concern for the welfare and health of the participant. In the playful condition, the robot repeatedly talked about having fun with the participant and told a joke. At the same time, we designed the dialogue so that the robot would seem equally humanlike and competent in both conditions. Over a period of about 20 minutes, the robot led the participants in a series of breathing and stretching routines. The routines got progressively more difficult. After the last routine, the robot asked participants to make up their own routine and do it as long as they could.

Here are examples from the serious versus playful scripts.

<u>Playful</u>
R: Do you like to exercise?
P: Kind of.
R. That's ok. These are fun-you'll love them.
Let's start. I want you to breathe to warm up. Do you know how to breathe?
P: Yes
R: Ha ha ha! I hope so. Ready to start?
R: Close your eyes.
R: Relax. Breathe in.
R: Don't forget to breathe out. I don't want you to pass out!

<u>Serious</u>
R: Do you exercise?
P: Sort of.
R: It is very important to your health.
R: I would like to have you do some exercises now. Would that be okay?
P: Sure.
R: Good, try to do everything that I say as best you can.
R: Let's start with a breathing exercise. Are you ready?
P: Yes.
R: Close your eyes.
R: Relax. Breathe in.
R: Breathe out. Are you feeling relaxed?

Measures

Measures of the mental model used the Big Five Inventory, the intellectual evaluation scale, and humanlikeness items. They were taken twice, once after the first series of exercises and then at the end of the study, after the robot asked the participant to make up an original exercise and do it as long as possible.

In this study, as compared with our earlier studies, the measures were implemented on a computer, and participants had to rate every trait. To measure mental model richness, we measured participants' reaction times in making ratings. Quicker ratings indicate a full mental model [7, 9]. The behavioral measure of cooperation was the time in seconds that the participant exercised when the robot asked the participant to make up a routine and perform it as long as possible.

Results

Participants rated the playful and serious robot as equally humanlike and intelligent. They also generated equally rich mental models in the two conditions, as measured by their reaction time in making personality ratings of the robot. The fastest times (most certain mental models) were achieved when participants rated the robot's agreeableness and (lack of) neuroticism; the slowest times, when they rated its openness to experience ($F [4, 152] = 6.2, p < .001$).

As we predicted, the content of personality ratings differed across the two conditions of the experiment. When participants interacted with the playful robot, they generally rated it more positively across all personality traits than did those who interacted with the serious robot ($F [1, 38) = 4.4, p < .05$). The interaction effect ($p < .05$) and contrasts show that these differences were especially apparent on the traits of extraversion and openness, where the playful robot was

[1] This study also included a manipulation and measures of participants' mood. Space precludes a complete description here, but main effects remain the same.

viewed as significantly more extraverted (t [2, 152] = 3.1, p < .01) and open to experience (t [2, 152] = 1.96, p < .05).

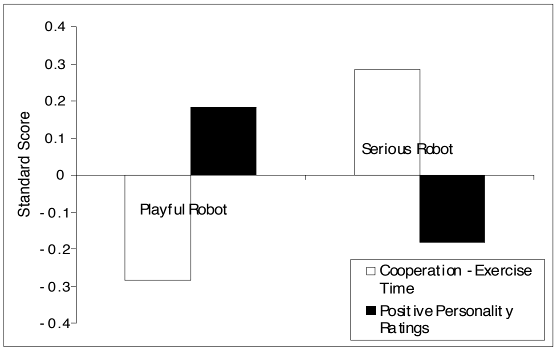

Figure 6. Personality ratings and cooperation. (Scores are standardized to permit comparison of the trends.)

Our predictions that the playful robot would elicit the most cooperation were not confirmed. One participant who interacted with the serious robot did not comply at all, but the other 19 participants who interacted with the serious robot exercised longer than the 20 who interacted with the playful robot. (F [1, 37] 6.2, p = .01). Including everyone, average exercise was 85 seconds with the serious robot versus 24.7 seconds with the playful robot. (See Figure 6.)

These results present a paradox. Why did participants not cooperate more with the robot they rated more positively, who seemed more fun and entertaining? Some explanations that fit our data include the following:

1. The serious robot may be perceived as caring about what a person does.
2. The serious robot may be more credible--more convincing that cooperation is important or exercise is healthy.
3. The serious robot may be seen as more likely to disapprove if the person does not exercise.

These possibilities are reminiscent of the dictum given new teachers, "Don't smile until Christmas!" (Sproull, personal communication). That is, perhaps the most effective robotic assistant requiring effort and cooperation from users does not have to be especially likeable or entertaining. Perhaps it is more important to create a serious or caring personality. Our current data certainly point to needed research on this question.

SUMMARY OF RESULTS AND DESIGN IMPLICATIONS

On first meeting, people have a comparatively sparse anthropomorphic mental model of an interactive robot. That is, they have no difficulty making mechanistic ratings but are slower and less willing to rate the robot on certain anthropomorphic dimensions of personality. Changes in the details of a robot's appearance and dialogue can enrich and redirect these mental models. For example, in one of our studies, adding hardware to the robot changed how reliable and powerful the robot seemed.

In our earlier studies, we noticed the robot's dialogue seemed to influence participants' mental models more than did differences in its appearance. Participants were significantly more likely to rate the robot's personality traits on dimensions the robot actually displayed in its dialogue (such as extraversion) than on traits the robot did not display. When we manipulated the robot's dialogue in study 5, we found that participants' mental models reflected the general personality given off by the robot. They perceived the playful robot as more extraverted and somewhat more open to experience than they did the serious robot, but they complied more with the serious robot.

We want to be appropriately cautious in making design recommendations based on this initial research. For example, perhaps, in our studies, the robot's appearance had comparatively small effects because we did not vary important aspects of robotic appearance, such as gender or age. We also have not yet examined how either robotic appearance or dialogue should be matched with the tasks the robot is to do. A robot's appearance and dialogue will need to be suited to its tasks and users, and to their context. However, we do have some evidence that simply creating a charming humanoid personality will not necessarily engender the best cooperation with a robotic assistant. In our study, the serious as compared with the playful robot elicited more cooperation. Perhaps a robot must change its personality to fit the task or indeed, the mood of the user, a demanding design requirement. Moreover, the irony of building a robot with a changeable personality is that this is likely to lead to an even more anthropomorphic mental model.

In this research, we have developed some prototype measurements for evaluating people's mental models of autonomous, interactive robots, and of their behavioral responses to these robots. We have demonstrated the usefulness of a tool kit of measurements by showing, in the experiment with the robotic assistant, that neither ratings nor behavioral observation alone would have been sufficient to describe participants' responses to that robot. The measures we have developed consist of (a) scales for rating anthropomorphic and mechanistic dimensions of people's mental model of a robot (b) measures of mental model richness or certainty, and (c) measures of compliance with a robot's requests. In future research, we plan to test the validity and reliability of these measures further, especially as they apply to interactions with different robots and across time. Meanwhile, we will offer our scales to other researchers on our website [deleted to maintain anonymity; to be added if paper is accepted].

CONCLUSION

As robotic assistants become more technologically advanced, they will require more interaction with operators, clients, and workers. A principled understanding of the cognitive and social nature of such interactions could significantly strengthen the field of robotics, open the way to more successful development of personal service robotics, and aid our understanding of the social and organizational impact of these robotic assistants.

ACKNOWLEDGMENTS

This research was supported by grants from the National Science Foundation, # 0085796 & #1120180. We especially thank Allyson Pottmeyer, who helped conduct the experiments, and Sebastian Thrun, Nick Roy, Joelle Pineau, Jamie Schulte, Mike Montemerlo, Greg Baltus, and Greg Armstrong from the Nursebot project. Esther Chen, Mike Schneider, Aaron Spaulding, Jack Zaientz , masters students in the HCII program, conducted the user test.

REFERENCES

1. Bates, J. (1994). The role of emotion in believable agents. *Communications of the ACM*, *37*, 122-125.
2. Black, A. W., & Eskenazi, M. (2001). A study of speech over the telephone and aging. Unpublished ms., LTI, Carnegie Mellon University, Pittsburgh, PA 15213.
3. Breazeal, C. (2000). Sociable machines: Expressive social exchange between humans and robots. Doctoral dissertation, Department of Electrical Engineering and Computer Science. MIT.
4. Caporael, L. R., & Heyes, C. M. (1996). Why anthropomorphize? Folk psychology and other stories (pp. 59-73). In R. W. Mitchell, N. S. Thompson, & H. Lyn Miles (Eds.), *Anthropomorphism, anecdotes, and animals*. Albany, NY: State University of New York Press.
5. Cialdini, R.B., Wosinska, W., Barrett, D.W., Butner, J., & Gornik-Durose, M. (1999). Compliance with a request in two cultures: The differential influence of social proof and commitment/consistency on collectivists and individualists. *Personality & Social Psychology Bulletin, 25*, 1242-1253.
6. Doyle, M. & Straus D. (1993) How to make meetings work: The new interaction method. Berkley Pub. Group.
7. Fabes, R. A., & Martin, C. L. (1991). Gender and age stereotypes of emotionality. *Personality and Social Psychology Bulletin*, *17*, 532-540.
8. Gilbert, D. K., & Rogers, W. A. (1999). Age-related differences in the acquisition, utilization, and extension of a spatial mental model. *Journals of Gerontology*: *Series B*: *Psychological Sciences & Social Sciences*. *54B* (4), 246-255.
9. Gill, M. J., Swann, W. B., & Silvera, D. H. (1998). On the genesis of confidence. *Journal of Personality and Social Psychology*, *75*, 1101-1114.
10. Gosling, S. D., & John, O. P. (1999). Personality dimensions in nonhuman animals: A cross-species review. *Current Directions in Psychological Science*, *8*, 69-75.
11. Hebb, D. O. (1946). Emotion in man and animal: an analysis of the intuitive processes of recognition. *Psychological Review*, 53, 88-106.
12. Heider, F. & Simmel, M. (1944). An experimental study of apparent behavior. *American Journal of Psychology*, *57*, 243-259.
13. Hintzman, D. (1986). Scheme abstraction in a multiple-trace memory model. *Psychological Review*, *93*, 441-428.
14. John, O. P., Donahue, E. M., & Kentle, R. (1991). The Big-Five Inventory -- Versions 4a and 54. Technical Report, Institute of Personality and Social Research, University of California, Berkeley, CA.
15. Kiesler, S., Sproull, L. & Waters, K. (1995). A prisoner's dilemma experiment on cooperation with people and human-like computers. *Journal of Personality and Social Psychology*, *79*, 47-65.
16. Linville, P. W., Fischer, G. W., & Salovey, P. (1989). Perceived distributions of the characteristics of in-group and out-group members: Empirical evidence and a computer simulation. *Journal of Personality and Social Psychology*, *57*, 165-188.
17. Mateas, M. (1999). An Oz-centric review of interactive drama and believable agents (pp. 297-398). In M. J. Wooldridge & M. Veloso (Eds.), *Artificial intelligence today*: *Recent trends and developments*. NY: Springer.
18. Nass, C. & Moon, Y. (2000). Machines and mindlessness: Social responses to computers. *Journal of Social Issues*, *56*, 81-103
19. Parise, S., Kiesler, S., Sproull, L., & Waters, K. (1999). Cooperating with life-like interface agents. *Computers in Human Behavior*, *15*, 123-142.
20. Rousseau, D. & Hayes-Roth, B. (1997, Sept.) Interacting with personality-rich characters. Technical Report No. KSL 97-06. Knowledge Systems Laboratory, Department of Computer Science, Stanford University, Stanford, CA. 94305.
21. Stapel, D. A., & Schwarz, N. (1998). Similarities and differences between trait concept and expectancy priming: What matters is whether the target stimulus is ambiguous or mixed. *Journal of Experimental Social Psychology*, *34*, 229-246.
22. Thrun, S., Beetz, M, Bennewitz, M., Burgard, W., Cremers, A. B., Dellaert, F., Fox, D., Haehnel, D., Rosenberg, C., Roy, N., Schulte, J., & Schulz, D. (2000). Probabilistic algorithms and the interactive museum tour-guide robot, Minerva. *International Journal of Robotics Research*, *19*, 972-999.
23. Warner, R.M. and Sugarman, D. B. (1986). Attributions of personality based on physical appearance, speech, and handwriting. *Journal of Personality and Social Psychology*, *50*, 792-799.