Proceedings of the 2005 IEEE
International Conference on Robotics and Automation
Barcelona, Spain, April 2005

Human Mental Models of Humanoid Robots*

Sau-lai Lee
Human Computer Interaction Institute
Carnegie Mellon University
5000 Forbes, Pittsburgh, PA 15232, USA
slleeh@hkusua.hku.hk

Sara Kiesler
Human Computer Interaction Institute
Carnegie Mellon University
5000 Forbes, Pittsburgh, PA 15232, USA
kiesler@cs.cmu.edu

Ivy Yee-man Lau
Department of Psychology
The University of Hong Kong
Pokfulam, Hong Kong
ilau@hkusua.hku.hk

Chi-Yue Chiu
Department of Psychology
The University of Illinois at Urbana-Champaign
603 Daniel Street, Champaign IL 61820, USA
kiesler@cs.cmu.edu

***Abstract* – Effective communication between a person and a robot may depend on whether there exists a common ground of understanding between the two. In two experiments modelled after human-human studies we examined how people form a mental model of a robot's factual knowledge. Participants estimated the robot's knowledge by extrapolating from their own knowledge and from information about the robot's origin and language. These results suggest that designers of humanoid robots must attend not only to the social cues that robots emit but also to the information people use to create mental models of a robot.**

Index Terms* – *human-robot interaction, social robots, humanoids, perception, dialogue

INTRODUCTION

Because people are social animals, robots that interact with people may be more effective communicators if they hold a correct theory of people's social expectations. Do people's mental models create an expectation that a robot knows what people know and do what people do? If so, does this similarity exist in all situations?

Participants in human experiments sometimes interact with desktop computer applications as though they were interacting with people [1-3]. In these studies, people appear to apply well-learned conventional social schemas (such as gender stereotypes) and norms (such as reciprocity) when they respond to the interactive system.

Social psychological research suggests at least two plausible theoretical explanations for people's apparent social responses to computer systems. One explanation is that people respond automatically to the social cues emitted by the system, and use these cues mindlessly (that is, without thoughtful mental processing); they simply apply stereotypes and heuristics, and enact social habits [4]. If this explanation is correct, people may respond automatically to social cues emitted by a robot, and apply human-human social schemas and norms to these interactions.

An alternative explanation of people's observed social responses to interactive systems is that this behavior is partly determined by their specific mental model about how and why systems behave as they do. If a system looks and behaves much like a human being (e.g., a humanoid robot emits a human's voice), their mental model of the system's behavior may approach their mental model of humans, but this model may differ in important respects from their models of humans [5]. For instance, in a previous study [6], participants played a Prisoner's Dilemma game involving real money with a real person or a computer agent. When the agent looked like a person, people cooperated with the agent at the same level as they did with the real person. When the agent looked like a dog, cooperation declined markedly, except in dog owners. A post-test survey of the participants suggested that participants who owned dogs had a mental model of the dog agent as cooperative whereas nonowners did not. These data suggest that mental models moderate people's responses to interactive systems.

The major difference between the two explanations is that the former implies people do not hold a theory or model about how or why robots behave. The latter explanation presumes people do hold such a theory. When the theory people hold for a robot is similar to their theories about people, they will interact with a robot and a person similarly, but if the theory people hold for a robot is dissimilar to their theories about people, then they will interact with a robot and a person differently. The first explanation also implies cross-task and cross-situation consistency if social cues are similar because these cues are used to generate the same social responses to a person and to a robot. The second explanation predicts task-specific and situation-specific interaction patterns, in which people's responses to a robot depend on their mental model of the robot in the given task and social situation. People might have similar mental models of a person and a robot in one task domain such as mathematical computation, but different mental models of a person and a robot in another task domain such as learning about landmarks.

To test which of these explanations is more valid, we conducted two controlled experiments in human-robot knowledge estimation. We asked participants to interact with a robot for a short time in a task domain that has been well established in social psychological research. Participants were told about a robot's origin, and then were asked to estimate its knowledge of landmarks in two locations. Using this approach, we were able to examine the existence and nature of people's mental models of robots.

The experiments we conducted have significance for the design of robots and human-robot interfaces. If the first explanation (automatic response to social cues) is valid, then design should focus primarily on identifying the social cues

* This work is supported by NSF Grant #IIS-0121426.

that robots should emit to elicit desired social responses from people. If the second explanation (mental models direct responses) is valid, then designers also must attend to the information people use to create mental models of a robot.

RELATED WORK

A. Socially Interactive Systems

In addition to the aforementioned work by Nass and his colleagues on desktop computers that emit social cues, considerable work has gone into the creation of social agents and characters that appear on computer displays, e.g., [7-9]. In the last decade, researchers have developed physically embodied mobile robots, such as robotic tour guides, that are meant to interact socially with people [10,11]. Minerva used reinforcement learning to adapt appropriately [12]. Kismet also is a robot whose purpose is to interact with people socially, and was developed on the model of an infant human [13]. Kismet emits emotional and social behavior to engage people. Vikia [11] and Valerie [14] are mobile robots that are designed for social interaction. Each of these robots uses social cues in speech and movement to create social responses among people. In an experiment, Bruce et al. [11] found that when Vikia had a simulated face and turned toward people passing by, passersby were more likely to respond positively to the robot. Another experiment on people's social responses to robots was performed by Goetz, Kiesler, and Powers [15]. They discovered that people cooperated more with a robot whose social behavior was matched appropriately with a task, e.g., cheerful behavior when the task was fun and serious behavior when the task was taxing.

Virtually all of this prior work has focused primarily on how the robot and its behavior can be designed (or can learn) to emit appropriate social cues and behavior. The current work focuses instead on how information emitted by a robot may create specific mental models of the robot in people who are to interact with it. To understand this problem, we apply previous social psychological research on human-human communication.

B. Human-Human Communication

How people form mental models of others is a complex question addressed in fields ranging from neuroscience to developmental psychology [16, 17]. We are interested here in one aspect of the mental models people hold of robots, that is, their estimates of the robot's knowledge. Knowledge estimation is a fundamental process in social interaction. All social interaction requires people to exchange information, e.g., their names, their goals, their emotions, etc. To exchange information successfully, people estimate what their shared common knowledge is and formulate their messages in respect to this shared knowledge [17]. For example, when strangers ask us for directions to a local restaurant, we estimate or determine where the strangers come from. If we perceive them to live in the local area, we also infer they know the names of local landmarks, and we use these names to tell them about the route to the restaurant. If we think they are not local, we will not use the names of local landmarks in referring to the route.

To estimate their common ground, communicators must go through a knowledge estimation process. Clark and his associates, e.g., [18], proposed that people used observable physical and linguistic cues to infer their common ground knowledge, as well as information they have about one another's group memberships, educational background, or professional identities. People are highly accurate in their estimates of the distribution of mundane knowledge in a particular population. For example, students were able to estimate the proportion of other students who knew the names of public figures [19] and landmarks [20] and the proportion of students who endorsed a particular set of values or experienced certain emotions [21]. Research also has shown that people's estimates of others' knowledge significantly affect how they communicate with those people. Thus, when participants were asked to describe public figures to another person, they provided descriptive information in inverse proportion to their estimates that the other person could identify the public figure [19].

This work points to the possibility that when people interact with social robots, their behavior will be influenced by their estimates of the robot's knowledge base. For instance, if people need to send a robot to a location and they assume the robot is familiar with the terrain, this knowledge should cause them to (a) use local landmarks to direct the robot, and (b) reduce the amount of information they give the robot (because they assume the robot already "knows" the area). If people are unfamiliar with the robot, how would they make these estimates? The previous work in social cues suggests that physical, linguistic, and social context cues will guide these estimates. Even the robot's origin, e.g., whether it is made in America or Asia, might be used as a cue to guide knowledge estimations. Thus, an American-made, English-speaking robot would be assumed to know better where the Empire State building is than a Hong Kong-made, Cantonese-speaking robot. The same process might be expected to affect not just people's estimates of the robot's knowledge of factual information, but also its beliefs or social preferences.

METHOD

We conducted two experiments to test the hypothesis that individuals' representation of a robot's knowledge would change when the origin of the robot changed. Chinese participants observed a robot interacting with the experimenter. Half of the participants saw the robot speak Cantonese with the experimenter (who was Chinese) and were told the robot was built at a robotics institute in Hong Kong. The other half of the participants saw the robot speaking English with the experimenter, and were told the robot was built at a robotics institute in New York. Then all participants saw photos of well-known and obscure tourist landmarks in Hong Kong and New York. They were asked to estimate the likelihood the robot could identify these landmarks. We compared participants' estimations of the robot's knowledge when the robot originated either in Hong Kong or New York.

We hypothesized that the origin of the robot and language it used would create different mental models of the robot in the minds of participants such that participants would believe the robot built in Hong Kong had knowledge of Hong Kong tourist landmarks, and that the robot built in New York had knowledge of New York tourist landmarks. We also expected participants to infer that both robots

would have greater knowledge of famous landmarks than obscure landmarks.

A. Participants

In Experiment 1, 60 Hong Kong students (19 males, 41 females; average age 21.15) from the University of Hong Kong participated in this study to fulfil part of a course's requirement.

In Experiment 2, 48 participants, 15 male and 33 female, average age 21.35, from the University of Hong Kong participated in the study. All were native Chinese and they had resided in Hong Kong for average 20.22 years. They received US$6 as payment.

B. Procedure

The stimuli and procedure used in this study were adapted from Lee & Chiu [22]. Lee and Chiu presented photographs of 14 landmarks to Hong Kong undergraduates and asked them to estimate the likelihood that the landmarks could be identified by Hong Kong undergraduates or undergraduates from New York. Participants saw landmarks that were famous and judged them to be familiar to everyone (e.g., the Statue of Liberty and the Great Wall of China). Other landmarks were thought to be more familiar to those living in Hong Kong (e.g., Hong Kong Cultural Center) or to those living in New York (e.g. Lincoln Center). Still other landmarks were judged unfamiliar to both Hong Kongers and New Yorkers (e.g., Kwoloon Wall City Park and the Dakota). In the Lee and Chiu study, students could accurately gauge others' knowledge of landmarks. Furthermore, their estimates influenced how they communicated when they were asked to describe the landmarks to another person. For instance, if they thought the person already knew a landmark, they spent less time describing the landmark to him or her.

In the current experiments, we asked participants to estimate the likelihood that a robot made in New York or Hong Kong would know and recognize landmarks in these cities. Half of the participants (HK condition) were told that the robot was built at a robotics institute at the Hong Kong University of Science & Technology and the other half of the participants (US condition) were told that the robot was created at a robotics institute in a university in the United States (Columbia University). Participants were shown pictures of these universities.

Fig. 1. Robot viewed in experiment.

All further instructions and stimuli were presented on a Powerbook G3 computer using the program, Power Laboratory. Participants were told that the aim of the research was to investigate how people communicate with robots. They were told they would make some judgments of a robot. The robot, they were told, was equipped with various speech recognition and speech production functions. It could understand English, Cantonese, and 16 other European and Asian languages. It could answer questions posed in speech or typing. We said field studies had demonstrated that the robot was effective in encoding and decoding different human languages.

In the US condition, participants were shown a video of the Pearl robot ambulating, and then approaching and interacting with the experimenter [23]. The robot and the experimenter, who could be identified as Chinese (like the participants), interacted with one another in English. In the video, the experimenter was seated with her back facing the camera. The script was tailored in such a way that it was synchronized with the lip movements of the robot. Participants in the HK condition received the same set of instructions except that in the video, the experimenter and the robot interacted in Cantonese, a dialect commonly used in Hong Kong. The robot's English speech synthesis was implemented using Cepstral's Theta (www.cepstral.com) and the Cantonese speech synthesis was implemented using CUTtalk (http://dsp.ee.cuhk.edu.hk/speech/cutalk/).

Participants in both conditions then completed the knowledge estimation task. First they viewed the set of 14 landmarks once. Next they were asked to view the landmarks one by one, and identify the landmarks themselves. Next they were asked to estimate the likelihood using a rating scale from 0% likelihood to 100% likelihood that the robot could identify each landmark. The order of presentation of the landmarks was randomised for each participant.

Fig. 2. Experimenter with robot, as seen by participants.

The knowledge estimation procedure in Experiment 2 was a replication of that in Experiment 1 with one exception. To avoid the possibility that some students in Hong Kong might not recognize Columbia University, we changed the identity of the U. S. university from Columbia University to New York University.

To rule out the alternative explanation that differences in estimations of knowledge of the robot were due to differences in the robot's perceived technical sophistication, we also asked the participants to rate the performance of the robot on three dimensions: its understanding of human

speech, its ability to talk, and its ability to communicate with people. Participants judged the robot's performance on three 7-point rating scales from 1 (poor) to 7 (excellent). Participants in Experiment 1 rated the robot's speech production and communication with humans similarly in the two conditions, but participants in the HK condition rated the robot's recognition of human speech more highly than did the participants in the US condition ($t(28)=-3.24$, $p<.05$). Participants in Experiment 2 did not rate the robots differently in any of the three dimensions. Because the knowledge estimation results of Experiment 1 and 2 were identical, we have some assurance that differences in knowledge estimates for the robots built in the two countries were not caused by differential perceptions of the robots' technical sophistication.

RESULTS

To recap, participants saw four groups of landmarks: landmarks familiar to people from both cultures, landmarks familiar to people who live in the U.S., landmarks familiar to people who live in Hong Kong, and landmarks unfamiliar in both cultures. For each participant, we averaged their estimations for each group of landmarks to create four average scores. Using the MANOVA technique, we tested statistically whether participants' estimations of the robot's knowledge was affected by the country of origin of the robot (between subjects) and the familiarity of the landmarks in a culture (within subjects). The analysis was a 2 (US condition versus HK condition) X 2 (Familiar versus Unfamiliar to Hong Konger) X 2 (Familiar versus Unfamiliar to New Yorker) MANOVA using each participant's four average scores.

A. Experiment 1 Results

The results of Experiment 1 showed first that participants extrapolated from their knowledge of people to estimate the robot's knowledge. Landmarks thought to be familiar to people living in Hong Kong were estimated to have an average 83% likelihood of being recognized by the robot as compared with just 48% likelihood if the landmarks were unfamiliar to people living in Hong Kong ($F [1, 28] = 132$, $p < .05$). Likewise, landmarks thought to be familiar to people living in New York were estimated to have an average 76% likelihood of being recognized by the robot as compared with just 55% likelihood if the landmarks were unfamiliar to people living in New York ($F [1, 28] = 61$, $p < .05$). Thus familiar landmarks were estimated to be more likely to be known by the robot than unfamiliar landmarks, regardless of where it was created.

A second result was a Condition X Familiar versus Unfamiliar to New Yorker interaction ($F [1,28] = 17$, $p < .05$). When participants were told that the robot was made in New York, they estimated the robot to be on average 77% likely to know the landmarks that were familiar to New Yorkers but only 46% likely to know landmarks that were unfamiliar to New Yorkers. By contrast, when participants were told that the robot was made in Hong Kong, they made no such differentiation (76% for landmarks familiar to New Yorkers versus 63% for landmarks unfamiliar to New Yorkers).

B. Experiment 2 Results

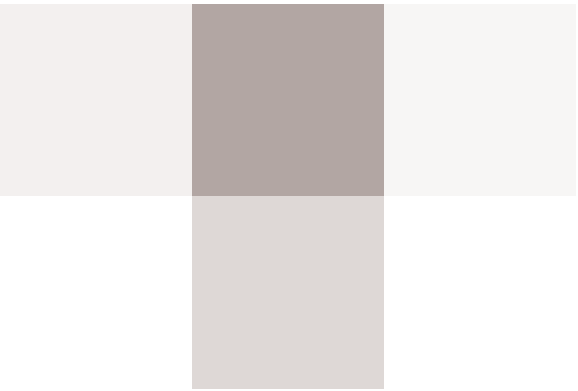
The results of Experiment 2 were similar to those of Experiment 1. First, participants thought the robot was more likely to identify landmarks that were familiar to people living in Hong Kong ($F [1, 41] = 110$, $p < .0001$ and landmarks familiar to New Yorkers, ($F [1, 41] = 58$, $p < .0001$. Also, there was a significant Condition X Familiar versus Unfamiliar to New Yorker interaction ($F [1, 41] = 9$, $p < .01$. When the participants were told that the robot was made in New York, they estimated the robot to be 80% likely to know the landmarks that were familiar to New Yorkers and just 61% likely to know the landmarks that were unfamiliar New Yorkers ($t [20] = 6.6$, $p < .05$). When the participants were told that the robot was made in Hong Kong, they also differentiated their estimates, but the difference was smaller than that found in the US condition (80% for familiar landmarks and 71% for unfamiliar landmarks, $t [22] = 7.1$, $p < .05$).

In sum, participants estimated the knowledge of the robot based on what they knew about people. They expected the robots to know more of the landmarks that were famous in both countries and less likely to know the landmarks that were unfamiliar to people in both countries. Also, the origin of the robot influenced their estimations. An American robot made in New York was perceived as more likely to know famous New York landmarks than obscure New York landmarks. A Chinese robot made in Hong Kong was perceived (significantly so only in Experiment 2) as more likely to know famous Hong Kong landmarks than obscure Hong Kong landmarks.

TABLE I
MEAN (SD) ESTIMATES OF A ROBOT'S KNOWLEDGE OF LANDMARKS IN HONG KONG AND NEW YORK[a]

Landmarks	Likelihood a robot created in New York would know the landmark (NY condition)	Likelihood a robot created in Hong Kong would know the landmark (HK condition)
Experiment 1		
Familiar to people in Hong Kong and New York	92%	89%
Familiar only to people in Hong Kong	64%	83%
Familiar only to people in New York	58%	57%
Unfamiliar to people in Hong Kong and New York	34%	48%
Experiment 2		
Familiar to people in Hong Kong and New York	89%	89%
Familiar only to people in Hong Kong	77%	88%
Familiar only to people in New York	68%	69%
Unfamiliar to people in Hong Kong and New York	49%	56%

[a]Estimates varied from a 0 to 100% likelihood that the robot would know the landmark.



[6] S. Parise, S. Kiesler, L. Sproull, and K. Waters. Cooperating with life-like interface agents. *Computers in Human Behavior*, vol. 15, pp. 123-142, 1999.

[7] J. Bates, "The role of emotion in believable agents," *Communications of the ACM*, vol. 37, no. 7, pp. 122-125, 1994.

[8] J. Cassell, T. Bickmore, H. Vilhjlmsson, and H. Yan, "More than just a pretty face: Affordances of embodiment," *Proceedings of the 2000 International Conference on Intelligent User Interfaces*, New Orleans, 2000.

[9] T. Koda and P. Maes, "Agents with faces: The effect of personification," *Proceedings of the 5th IEEE International Workshop on Robot and Human Communication* (ROMAN 96), pp. 189-194, 1996.

[10] C. Breazeal and B. Scassellati, "How to build robots that make friends and influence people," *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems,* Knyoju, Japan, 1999.

[11] A. Bruce, I. Nourbakhsh, and R. Simmons, "The role of expressiveness and attention in human-robot interaction," ICRA, Washington, D. C., Mary 2002.

[12] S. Thrun, M. Beetz, M. Bennewitz, W. Burgard, A. B. Cremers, F. Dellaert, D. fox, D. Haehnel, C. Rosenberg, N. Roy, J. Schulte, and D. Schulz, "Probabilistic algorithms and the interactive museum tour-guide robot Minerva, International Journal of Robotics Research, vol. 19, no. 11, pp. 972-999, 2000.

[13] C. Breazeal (Ferrell) and J. Velasquez, J. "Toward teaching a robot `infant' using emotive communication acts," *Proceedings of 1998 Simulation of Adaptive Behavior (SAB98) Workshop on Socially Situated Intelligence*, Zurich, Switzerland, 1998.

[14] http://www.roboceptionist.com/

[15] J. Goetz, S. Kiesler, and A. Powers, "Matching robot appearance and behavior to tasks to improve human-robot cooperation." Robot and Human Interactive Communication, 2003. Proceedings of ROMAN 2003 (pp. 55-60). The 12th IEEE International Workshop on, Vol., IXX Oct. 31-Nov. 2, 2003, Milbrae, CA.

[16] D. J. Povinelli and J. M. Bering, "The mentality of apes revisited," *Current Directions in Psychological Science*, vol. 11, no. 4, pp. 115-119, August 2002.

[17] R. S. Nickerson, "How we know—and sometimes misjudge—what others know: Imputing one's own knowledge to others," *Psychological Bulletin*, vol. 125, no. 6, pp. 737-759, 1999.

[18] E. A. Issacs and H. H. Clark, "References in conversation between experts and novices," *Journal of Experimental Psychology:General*, vol. 116, no. 1, pp. 26-37, 1987.

[19] S. Fussell and R. Krauss, "Coordination of knowledge in communication: Effects of speakers' assumptions about what others know." *Journal of Personality and Social Psychology*, vol. 62, pp. 378-391, 1992.

[20] I. Y-M. Lau, C. Chiu, and Y. Hong, Y. "I know what you know: Assumptions about others' knowledge and their effects on message construction." *Social Cognition*, vol.19, pp. 587-600, 2001.

[21] S-L. Lee & C. Chiu, "Judgmental accuracy: Effects of social projection and response typicality." Unpublished. Hong Kong University, HK., 2000.

[22] S-L. Lee & C. Chiu, "Communication and shared representation: The role of knowledge estimation." Unpublished. Hong Kong University, HK., 2003.

[23] www.cs.cmu.edu/~nursebot/

[24] B. J. Scholl and P. D. Tremoulet, "Perceptual causality and animacy," Trends in Cognitive Science, vol. 4, pp. 200-309, 2000.

[25] B. Scassellati, "Theory of mind for a humanoid robot," Autonomous Robots, vol. 12, pp. 13-24, 2002.

[26] A. Powers, A. Kramer, S. Lim, J. Kuo, S-L. Lee, S. Kiesler, "Common Ground in Dialogue with a Gendered Humanoid Robot," unpublished.