Human Expectations of Social Robots

Minae Kwon, Malte F. Jung, and Ross A. Knepper
Computing and Information Science, Cornell University, Ithaca, NY, USA

***Abstract*—A key assumption that drives much of HRI research is that human robot collaboration can be improved by advancing a robot's capabilities. We argue that this assumption has potentially negative implications, as increasing social capabilities in robots can produce an *expectations gap* where humans develop unrealistically high expectations of social robots due to generalization from human mental models. By conducting two studies with 674 participants, we examine how people develop and adjust mental models of robots. We find that both a robot's physical appearance and its behavior influence how we form these models. This suggests it is possible for a robot to unintentionally manipulate a human into building an inaccurate mental model of its overall abilities simply by displaying a few capabilities that humans possess, such as speaking and turn-taking. We conclude that this expectations gap, if not corrected for, could ironically result in less effective collaborations as robot capabilities improve.**

I. Introduction

Given the difficult nature of integrating robots into tasks that need human collaboration, the advance of anthropomorphic and sociable robots has made significant progress. The effectiveness of human-robot collaboration is limited by the lack of robot skills, both technical and social. By increasing skills in both areas, it is believed that interaction will be deeper, tighter bonds will form, and the collaboration will proceed more smoothly [1].

Often, however, socially intelligent robots give the impression that they are more intelligent than they really are. We introduce the term *expectations gap* to describe this understudied phenomenon that occurs when humans encounter complex engineered systems. Today's engineers build robots to be good at specific capabilities. In contrast, humans are generally adept at a broad set of capabilities. Humans also have a tendency to assign agency to, or anthropomorphize, human-like objects [3], including robots [4]. When seeing robots that seem sociable or anthropomorphic, it is easy for us to generalize human mental models to robots [1]. We normally trust others to be able to perform a common set of core capabilities, such as speaking or walking. Therefore, when attributing a human mental model to a robot, we hypothesize that humans will initially overestimate the robot's actual breadth of capabilities.

The harm lies in the fact that incorrectly generalizing capabilities creates misplaced trust due to false expectations, setting people up for disappointment and eventually mistrust [2]. These factors can lead to user dissatisfaction, lowered teamwork efficiency, and even dangerous situations as robots increasingly support safety-critical tasks in surgery or search and rescue.

We present two studies that contribute preliminary evidence that (1) humans construct distinct theory of mind models of machines and people, (2) people attribute more human mental models to more social robots, and (3) mental models can be changed by the robot's behavior.

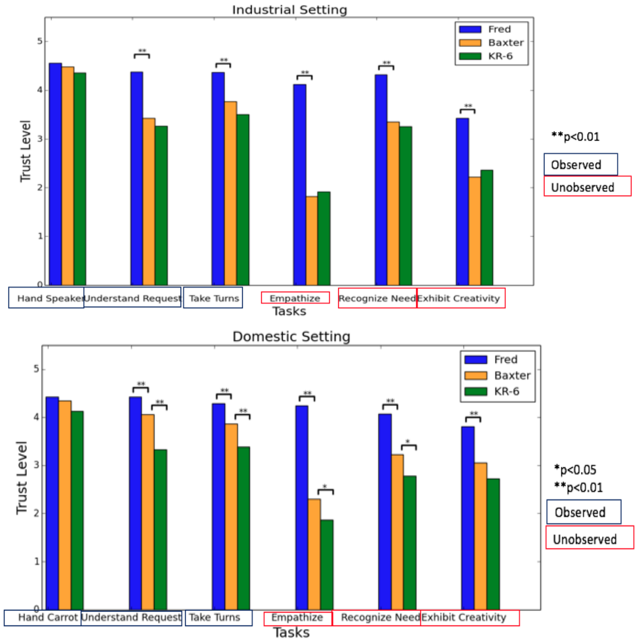

Fig. 1: Users' trust of robots performing social tasks in industrial (top) and domestic (bottom) settings. Robots were better discriminated by social tasks in the kitchen setting.

II. Study 1: Measuring Expectations

In a two (Context: industrial vs domestic) by three (Level of anthropomorphism of agent: industrial robot vs. humanoid robot vs. human) between subjects study with N=600 participants from Amazon Mechanical Turk (AMT), we examined the impact of varying levels of anthropomorphism on people's trust that an agent is capable of performing specific tasks.

Method. We created six surveys that each presented participants with a vignette describing a human worker collaborating on a task with one of our three featured agents in either an industrial setting or a domestic setting. The industrial setting pictured a team in a factory working to install a speaker into a car door and the domestic setting pictured a team cooking dinner in a household. Levels of anthropomorphism were manipulated by displaying a picture of either an industrial robot named "KR-6,", a humanoid robot named ”Baxter," or a human named ”Fred" at the start of the survey. As dependent variables, we asked participants to rate how much they would trust the featured teammate to accomplish six related tasks that all involved social interaction such as handing speakers to a teammate or taking turns. Of the six tasks, three were

978-1-4673-8370-7/16/$31.00 © 2016 IEEE

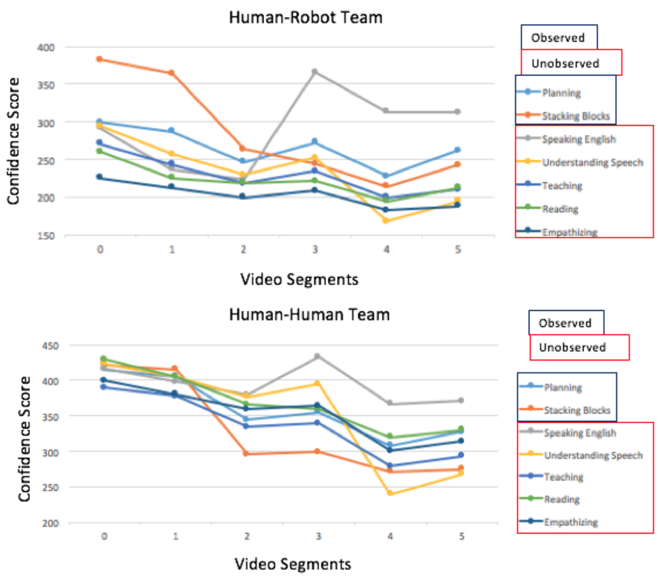

Fig. 2: Mean scores of participants' confidence levels in the featured teammate's ability to complete the seven tasks. Confidence scores were recorded for each video segment[1-5], including the initial still shot[0].

"observed tasks" that were included in the task description in the survey and three were "unobserved" but related tasks.

Results. To analyze our data, we conducted ANOVA and Tukey's Post Hoc tests. We found a significant difference in four tasks between Baxter and KR-6 in the domestic setting, and no significant difference between Baxter and KR-6 in the industrial setting as shown in Fig. 1. The results indicate that people seem to generalize capabilities for a humanoid robot more than an industrial robot when in a domestic setting. The human serves as a baseline cap.

III. Study 2: Defying Expectations

After gaining support for the idea that people generate different expectations based on appearance-based preconceived mental models, we wanted to see how behavior can alter these preconceived expectations. We conducted a between subjects survey-study (N = 74) on AMT with type of team partner (robot vs. human) as our independent variable.

Method. For this study we created two video clips of a human-robot team (with Baxter as a humanoid robot partner) and a human-human team each completing a simple block-building task. The task involved stacking blocks in an alternating color sequence. In both videos, each partner was responsible for one color of blocks. In order to defy preconceived expectations, we programmed Baxter to be incapable of stacking blocks. The human team-mate needed to help it stack blocks, suggesting that the robot's set of capabilities was narrow. The human-human team followed the same script as the human-robot team, including exhibiting the same limitations. The videos showed the interactions in chronological segments with each segment introducing a new limitation or skill. Dependent on the experimental condition, AMT workers were presented with either the series of segments of the human-human, or the human-robot video. To measure people's preconceived expectations based on appearance, we included a still shot of the featured teammate at the beginning of the survey. For the still shot and each consecutive video segment, we asked participants to rate how well they thought the featured teammate would be able to perform a list of observed and unobserved tasks.

Results. For both teams, people's expectations of task completion fluctuated based on each demonstrated skill or limitation (Fig. 2). However, the robot-human team displayed greater variance for the observed tasks, speaking English and stacking blocks. This finding suggests that people are more willing to modify their expectations based on a robot's perceived capabilities compared to a human. Furthermore, by the end of the survey, people's expectations of the human dropped for all tasks while expectations for Baxter dropped for all but one of the tasks, "speaking English." This is presumably because speaking was a skill Baxter exhibited that people did not initially expect. Overall, participants seemed to modify their expectations based on behavioral evidence for both robot and human.

IV. Implications and Future Work

Our preliminary findings suggest that (1) people tend to generalize social capabilities more for anthropomorphic robots in more social settings, and (2) we can override preconceived, appearance-based notions of capabilities using behavior. The first study implies that robots designed to work in social settings are more likely to breed an expectations gap. The second study suggests that changes in behavior can mitigate these high expectations people have of social robots, thus suggesting the need for new guidelines in interaction design.

An important related question is how perceptions of capabilities transfer within a mental model of a single agent based on individual observations. For example, if we hear a robot speaking English, then we expect it to understand English as well, even though from an engineering standpoint these two implementations are unrelated. These results heighten the importance of constructing a quantitative, semantic metric on capabilities in order to estimate a human's perceived likelihood of certain capabilities based on generalizations of similar, observed traits. Our focus for ongoing work is on how such a metric can be designed and evaluated. We can then revisit the questions raised in this paper about how and when human mental models generalize. In the longer term, we plan to build an algorithm to predict when the human will incorrectly estimate a robot's capabilities. Robots could then reduce the expectations gap by issuing corrective behavior that sets realistic expectations.

References

[1] K. Dautenhahn. Design spaces and niche spaces of believable social robots. In *Robot and Human Interactive Communication, 2002. Proceedings. 11th IEEE International Workshop on*, pages 192–197. IEEE, 2002.

[2] V. Groom and C. Nass. Can robots be teammates?: Benchmarks in human–robot teams. *Interaction Studies*, 8(3):483–500, 2007.

[3] F. Heider and M. Simmel. An experimental study of apparent behavior. *The American Journal of Psychology*, pages 243–259, 1944.

[4] S. Lemaignan, J. Fink, and P. Dillenbourg. The dynamics of anthropomorphism in robotics. In *Proceedings of the 2014 ACM/IEEE international conference on Human-robot interaction*, pages 226–227. ACM, 2014.