What Kind of Mind Do I Want in My Robot?

Developing a Measure of Desired Mental Capacities in Social Robots

Bertram F. Malle
Department of Cognitive, Linguistic, and Psychological Sciences
Brown University Providence, RI 02906
Email: bfmalle@brown.edu

Stuti Thapa Magar
Department of Cognitive, Linguistic, and Psychological Sciences
Brown University Providence, RI 02906
Email: stuti_thapa_magar@brown.edu

ABSTRACT

We took a first step in developing a measurement tool for the social and mental capacities that people desire in social robots. 291 respondents indicated the degree to which they would like to see 16 capacities in either a home robot, nursing robot, or military robot. Four orthogonal dimensions emerged: Social-Moral Skills, Autonomous Evaluation, Objective Reasoning, and Negative Feelings. These dimensions were robust across the three robot types and across two assessment years (2013 and 2016).

Keywords

Human-robot interaction; robot ethics; robot design; social robots

1. INTRODUCTION

When (not if) robots become widely available in society, numerous factors will determine people's acceptance of such robots. One key determinant will be people's expectations of such robots, and whether robots can meet those expectations. Previous work on expectations focused on appearance and task domains [1, 3]. However, many future robots will be interacting with humans, so those social robots will need to have social and mental capacities that people expect of their interaction partners. In the present project we therefore ask: *What kinds of social and mental capacities do people expect of robots?* We report first insights from an ongoing effort of developing a new measurement tool.

We operationalize "expectations" as people's desires for what social and mental capacities they want future robots to have. Because there is no prior work on such desires, we drew on literatures examining people's *beliefs* about robots' mental capacities [5], their moral competence [4], as well as their capacity for experiences and agency [2].

2. METHOD

We created an item pool of 16 potentially desired social-mental capacities in robots, informed by previous work on people's beliefs about similar capacities [2, 5] and HRI studies on morality and social relations. We included items that provided an array of social capacities (e.g., forming friendships, feeling attachment), affective capacities (having emotions, feeling stress), moral capacities (e.g., having values, engaging in moral deliberation, blaming/praising), autonomy (making independent choices), and transparency (explaining reasons for its actions). We omitted items that were too abstract (e.g., consciousness, personality), are undoubtedly necessary in robots (e.g., vision, hearing, memory, planning), or are inappropriate for robots (e.g., hunger).

HRI '17 Companion, March 6-9, 2017, Vienna, Austria
ACM 978-1-4503-4885-0/17/03.
http://dx.doi.org/10.1145/3029798.3038378

U.S. community participants from Amazon's Mechanical Turk crowdsourcing site completed an online experiment either in 2013 (n = 114) or 2016 (n = 177). People indicated how much they wanted to see each of 16 potential capacities in future robots, using 7-point rating scales from -3 (Very much NOT want) to +3 (Very much WANT). In each cohort, people were randomly assigned to consider either a military robot (that saves injured soldiers, defuses bombs, etc.), a nursing robot (that works in a hospital and performs medical care actions), or a home care robot (that cares for older adults or children and does household chores).

3. RESULTS

To detect the dimensional structure of people's desires we performed a Principal Component Analysis (Kaiser-Meyer-Olkin sampling adequacy = .90). We found a four-dimensional solution that explained 65% of the total variance. After orthogonal (Varimax) rotation, the first component was by far the strongest and had 8 items with substantial loadings ($\geq$ .60); it can be labeled Social-Moral Skills. The other three components, of nearly equal strength around 12% explained variance, each had 2 to 3 items with substantial loadings. We label them Autonomous Evaluation, Objective Reasoning, and Negative Feelings (see Table 2 for factor loadings). There were no substantial differences in dimensional structure when breaking the data by type of robot people considered or by cohort (2013, 2016).

Next we treated the four components as separate scales (αs = 0.60 to 0.90) and examined the profile of people's desires across the four capacities. First, people are right in the middle between wanting and not wanting robots to have Social-Moral Capacities (M = 0.10, SD = 1.56). The individual item means in Table 1 show that this mean conceals some variation—people do like to see empathy, goals, and values but not emotions in future robots. Second, people do not want robots to have Autonomous Evaluation capacities (M = -0.97, SD = 1.42) and clearly not Negative Affect (M = -1.62, SD = 1.61). But they very much want robots to have Objective Reasoning (M = 1.47, SD = 1.34).

Comparing the four subscales across the three robot use domains, we found no significant differences, $F(8, 566) = 1.24$, $p = .27$, $\eta^2 = .03$. Comparing across years, we found small differences, $F(4, 282) = 5.27$, $p < .001$, $\eta^2 = .07$. Specifically, the desire for Social-Moral Skills increased slightly from 2013 (M = -0.12) to 2016 (M = 0.25), $F(1, 285) = 3.85$, $p = .05$, $d = 0.24$; while the desire for Objective Reasoning decreased slightly from 2013 (M = 1.66) to 2016 (M = 1.35), $F(1, 285) = 3.83$, $p = .05$, $d = 0.24$.

Table 1: Mean ratings (-3 to +3) and principal component loadings of social-mental capacities that people desire in future robots

Mean Rating	Features	Social-Moral Skills	Autonomous Evaluation	Objective Reasoning	Negative Feelings
-0.11	*Forms friendships*	0.83			
0.96	*Empathizes with humans*	0.75			
0.63	*Has values*	0.74			
-0.26	*Feels attachment to certain people*	0.74	0.31		
0.09	*Praises moral*	0.69			
-0.82	*Has emotions*	0.65			0.39
-0.26	*Decisions based on moral deliberations*	0.63	0.46		
0.57	*Has goals*	0.60			
-1.34	*Blames immoral behavior*	0.32	0.71		
-1.50	*Likes some people, dislikes other people*	0.34	0.70		
-0.08	*Makes independent choices*		0.68	0.45	
1.95	*Thinks logically and rationally*			0.74	
1.58	*Explains the reasons for its own actions*			0.69	
0.88	*Feels heat, cold, pressure, etc.*			0.65	0.37
-1.38	*Feels pain*				0.87
-1.86	*Feels stress*				0.79
	Explained variance:	27.4%	12.9%	12.4%	12.0%

Notes: Mean ratings: Green cells are significantly (p <.01) above the zero mark (capacity is desired) and red cells are below the zero mark (capacity is undesired). Principal components analysis: Varimax-rotated solution (loadings < 0.3 not shown). When the gray-shaded item sets are treated as variables of four scales representing the four components, reliabilities (standardized Cronbach's α) are 0.90, 0.68, 0.60, 0.77, respectively.

4. DISCUSSION

We took a first step in developing a measure of people's desires for future robots' social-mental capacities. We found a four-dimensional structure that was robust across three use domains and two cohorts. The first component (Social-Moral Skills) accounted for twice as much variance as each of the remaining components, signaling the importance of moral capacities in future robots [5, 8]. However, only two to three items loaded on each of these remaining components. A next step in scale development is therefore to expand the item pool for these components and sharpen their meaning. For example, the second component has both an autonomy and an evaluative element; and the third component involves both sensing and reasoning. We also plan to expand the item base in other directions, such as to include communicative and social-cognitive capacities (theory of mind).

In the present data, some of the most wanted capacities were logical thinking, explaining the reasons for one's actions, and empathy; some of the least desired capacities were feeling stress and pain, having emotions, and evaluating others in blame or liking/disliking. It appears that people specifically reject robots with various forms of negative affect. Due to space constraints we were not able to report on a second part in the survey, which supports this interpretation. People rated whether they would want robots to have any of 23 specific emotions. The results were clear: People did not want robots to have any of the negative emotions and, among the positive ones, they only wanted compassion, gratitude, and happiness, primarily for home and nursing robots.

A final pattern to note is the consistent four-dimensional structure and consistent profile of preferences over the 3-year span, despite recent accelerated scientific and public discourse about AI and robotics. A lack of documented change is of course not necessarily evidence for an absence of change. Participants may have failed to observe or take part in the swell of both fear and excitement about AI and robotics that has emerged over the past years. However, our study population (Amazon Turk participants) is likely to have encountered news about AI and robotics; so our results may point to a stable set of expectations that demand further research, especially with samples from other cultures.

Our study is only the first step in a more extensive scale development project. With a reliable set of scales, we will be able to expand the types of robots to consider and assess changes in people's expectations of robots before and after brief or extended encounters. But even now, we can see these expectations are multi-dimensional and robust.

5. ACKNOWLEDGMENTS

This project was supported in part by a grant from the Office of Naval Research, No. N00014-13-1-0269. The opinions expressed here are our own and do not necessarily reflect the views of ONR.

6. REFERENCES

[1] Goetz, J. et al. 2003. Matching robot appearance and behavior to tasks to improve human-robot cooperation. *The 12th IEEE International Workshop on Robot and Human Interactive Communication* (Oct. 2003), 55–60.

[2] Gray, H.M. et al. 2007. Dimensions of mind perception. *Science*. 315, 5812 (Feb. 2007), 619–619.

[3] Haring, K.S. et al. 2016. How people perceive different robot types: A direct comparison of an android, humanoid, and non-biomimetic robot. *Proceedings of the 8th International Conference on Knowledge and Smart Technology (KST), 2016*. IEEE. 265–270.

[4] Malle, B.F. and Scheutz, M. 2014. Moral competence in social robots. *Proceedings of IEEE International Symposium on Ethics in Engineering, Science, and Technology, Ethics'2014*. IEEE. 30–35.

[5] Sytsma, J. 2014. The robots of the dawn of experimental philosophy. *Current controversies in experimental philosophy*. E. Machery & E. O'Neill, eds. Routledge. 48–64.

[6] Wallach, W. and Allen, C. 2008. *Moral machines: Teaching robots right from wrong*. Oxford University Press.