Social Robots Are Like Real People: First Impressions, Attributes, and Stereotyping of Social Robots

Byron Reeves, Jeff Hancock, and Xun Liu
Department of Communication, Stanford University

Research shows that particular social robots are effective in interactions with people, from teaching young children to supporting older adult's healthcare. There is little research, however, that examines how people form the interpersonal impressions of social robots that influence whether these interactions are effective or not. We report two studies that look at the differences between social robots with the goal of identifying those impressions. In Study 1, we examined 10 years of published articles about social robots and identified 342 robots for coding on a variety of attributes, from perceived gender to surface textures and their range of motion. We asked participants ($N = 4,415$) to evaluate each robot along these attributes. A complete catalog of the robots is posted here, and an overall description of the sample is included in the article. In Study 2, we asked people ($N = 3,920$) to evaluate the 342 robots and found that two impressions of the robots, competence and warmth, were as important for the evaluation of robots as they have been for perceptions of people in prior literature. Using the 21 attributes identified in Study 1, we found that the best predictors of robot competence were a robot's mobility (degrees of freedom) and surface textures, while age, the absence of mechanical features, and mobility were the best predictors of robot warmth. We end with comments about the design and success of social robots in assistive applications.

Keywords: social robots, human–robot interaction, social perceptions of technology

Supplemental materials: https://doi.org/10.1037/tmb0000018#supplemental-materials

Disclaimer: Interactive content is included in the online version of this article.
Contact Information: Correspondence concerning this article should be addressed to Byron Reeves, PhD, Department of Communication, Stanford University, Stanford, CA, USA. Email: reeves@stanford.edu

There have been over one thousand published scholarly articles about social robots in the academic literature, indicative of considerable interest as evidenced by new research labs, conferences, journals, TED talks, companies, patents, and products. The social robots of interest are autonomous physical embodiments (i.e., they exist in the real world not just on a screen), and they communicate with humans via social behaviors (e.g., speech, gestures, and movement) that mimic human interactions that are linked with particular social roles that a robot might play. The proposed roles for the robots vary, from teaching to childcare to toys to elder companionship, but the underlying rationale for their popularity is similar across contexts. Social robots can be designed with human-like features, textures, and movements, and programmed to behave in variety of social or helpful ways, enough so that the people who use them could interact via familiar and pleasurable and even emotional connections. Robots have the added advantage of lower cost, continuous availability, and constancy of responses compared to human counterparts. The potential for robots to improve human life across a range of capacities is high.

How might social robots achieve this potential? To date, there is limited conceptualization or generalized theory of robots as a *category of social actors*. Instead, research has favored the study of specific robots in unique contexts without substantial consideration of the generalizability of particular attributes to other robots. Examples of social robots, both in the marketplace and in academic research, have grown so fast that it is been difficult to collect and understand them as a category of technology. Three examples from the collection of robots described in this article highlight this issue. There is a 16-in. tall bear with a plush coating of fur, large friendly eyes, and a screen embedded in its chest; a printed-book sized cardboard box on tractor tires with two half-inch lenses for eyes; and a 4-ft tall bipedal creature with intricate mechanical and electrical

parts clearly visible on limbs and body but no facial features at all on a human sized head (see Figure 1).

These examples easily meet the description of social robots in the first paragraph, but they also have obviously quite different psychological prospects with respect to a social evaluation by humans. Consequently, we need research that both collects and describes social robot attributes across a large sample of robots and explores how people perceive the combinations of these attributes with respect to impressions people form about the robots as social actors. If we understand how people differentiate this diverse group of stimuli, and which evaluations are attributable to different robot features, then we would have a start on a basic understanding of human–robot interactions that could be used to study and design better human–robot interactions.

We draw on research from psychology that has examined how people perceive other people to explore whether those perceptions also apply to social robots. For example, research examining how people form first impressions and stereotype others suggests that we perceive people along two primary dimensions, warmth and competence. One prominent model, the stereotype content model (SCM, Fiske, 2018; Fiske et al., 2002) argues that we use these primary dimensions to stereotype people. Do we also evaluate social robots along warmth and competence dimensions (Russell & Fiske, 2008)? And do people stereotype the robots along these same dimensions?

The research reported here addresses these questions by examining what makes robots similar and different across a range of robot examples. In Study 1, we collect and describe a large sample of social robots on an array of attributes thought to be influential in social interaction. A goal of the study was to characterize the larger category of social robots on features that may determine how people interact with them. In Study 2, we evaluate the combination of possible descriptors of social robots that people use to differentiate the large sample of robots, concentrating on hypothesized similarities between humans and robots with respect to the social attributes of warmth and competence. Once confirmed, we then looked at which attributes of robots were the best predictors of the central perceptual dimensions of warmth and competence.

Background: Sampling and Describing Social Robots

Much of the past research has studied social robots by demonstrating that there are example single robots that can play specific roles, such as teaching, or accomplish particular tasks, such as

Figure 1
Examples of Social Robots

providing emotional support to an older adult, that are typically performed by humans. Research often creates experimental comparisons between robots and humans (Kanda et al., 2004); for example, comparing how autistic children learn a social behavior from a robot compared to an adult instructor (Cabibihan et al., 2013). Social robots often perform well, or at least well enough to be potentially valuable human surrogates. Other research tries to identify the features, appearances, mechanics, movements, and behaviors that will make robots successful for one particular task. This question is often studied by having the same robot perform or behave differently while interacting with people, highlighting different capabilities, or appearances during an interaction (Bruce et al., 2002; Kidd et al., 2006). For example, the same robot might move quickly or slowly during an interaction (Satake et al., 2009), speak confidently or hesitatingly about a topic (Lee et al., 2006), or look directly at or away from a human during interaction (Kozima et al., 2003). In addition to these controlled experiments, there are also numerous qualitative observations of human–robot interactions that provide rich descriptions of when and how interactions appear to succeed in specific contexts (e.g., Sabanovic et al., 2006).

Most research explores social robots *one robot at a time*, a strategy that limits research to the particular qualities of specific robots. This approach ignores important qualities that *differentiate* the robots and that might explain their success generally. There have been attempts to develop databases of social robots (Juarez et al., 2011; Kalegina et al., 2018; Phillips et al., 2018); however, they often focus on specific robot designs, such as rendered robot faces (Kalegina et al., 2018), or on specific characteristics such as anthropomorphism (Phillips et al., 2018).

There is value in continuing to collect multiple examples of social robots that capture the variance across an increasingly large number of robot designs. Social robots are a new category of interactive technology (e.g., like computer interfaces, virtual social agents, or virtual reality games). The boundaries of this category, however, are still fuzzy operationally given that new examples appear often. Given the heterogeneity of the robots, it is currently difficult to know with confidence how they will be perceived when studies choose only a single example of a diverse and large technology category. Similarly, single examples of robots make it impossible to compare attributes between different robots with respect to their importance for human perceptions.

In psychology generally, this is a problem of stimulus sampling. For example, if you are interested in how an individual's personality might influence impressions of that person in a relationship, you should not choose a single individual to represent all introverts and one other person to represent all extroverts. Too many other attributes of the two people chosen may influence the results. For example, the introvert may be physically attractive and the extrovert physically awkward; the introvert might be well spoken and the extrovert linguistically hesitant, and so on across many other common characteristics of people. It is critical to sample people of each personality, ensuring the distribution of nonrelevant attributes is equivalent across the stimulus examples for each category, just as would be true for samples of people that might respond to the stimuli. In psychological research, this is discussed as treating stimuli (i.e., social robots in the present case) as a separate source of random variance (Judd et al., 2012), just like the variance between people in an experiment. Research that ignores stimulus

variance cannot draw inferences, even if the sample of participants who participate in the research is otherwise adequate.

For social robots, the issue of stimulus sampling may be more important than it is generally in psychology, for two reasons. First, the stimuli (robots) are the entrée to research. A main purpose of the research is to know *which* robots will be more successful than others, at least as much as knowing how different groups of people will respond to particular ones. Consequently, as a research investment, sampling the variance between robots is critical, perhaps more so than sampling the variance among the people that respond to them. We note that this may not be the same research investment required for other areas of robotics. Testing the effectiveness of mechanics, materials, electronics, or software, for example, may be done adequately in a single robot body without jeopardizing conclusions about value across the category of robots.

Second, the variance between social robots is likely quite large because it is possible to build robots to extremes on multiple dimensions. Far beyond the range for humans, they can be tiny or huge, have no eyes or eyes as large as their face, have more or fewer than four limbs, possess animal or machine features, and all other sorts of humanly impossible appendages, sensors, screens, and shapes. Consequently, the large variance among social robots may require sample sizes *larger* than would be required to represent humans.

Study 1: Social Robots Collection in Published Research

Our first task was to assemble a large sample of social robots. There is no published review of the different examples of social robots that could guide research about how robots are different or whether any of their differences matter in human–robot interactions. In an attempt to sample as many current robots as possible, we reviewed the research on social robots over a 10-year period and cataloged the social robots in the published literature. A review of academic literature might miss robots that have been in the marketplace but never used in research; however, our goal was only to create a large sample of robots rather than to build a census of all that have been built for any reason.

Google Scholar was used to search the keyword combination of "social robot" for each year from 2005 to 2016. The complete list of studies was reduced to those conducted about human responses to the robots. Six research assistants reviewed each article to determine if it should be included in our final sample. Three inclusion criteria were used as follows: (a) a social robot must be named in the article as the focus of the research; (b) there must be a picture(s) of the robot in the article or in supplemental materials; and (c) the relationship of the social robot with people must be discussed and studied. This excluded articles that only referenced social robots by category, and studies that focused only on the technical features of robots.

All of the articles are referenced in the Stanford Social Robot Collection, a database that is publicly available (https://goo.gl/eejbV7), including citations, brief notes about the articles, the type of research (experiment, discussion, survey, and review), the name of the robots referenced in the article, and links to pictures of the robot if they were separate from the article. A more detailed compilation of the robots can also be found in the Stanford Social Robot Collection (https://goo.gl/Gqpzkx) including multiple perspectives of the robots where available.

Table 1
*Total Number of Articles and Inclusion-Satisfying Articles by Year*

Year	Total Google Scholar entries	Articles satisfying inclusion criteria
2005	120	35
2006	175	59
2007	221	56
2008	260	63
2009	363	106
2010	441	171
2011	542	148
2012	695	82
2013	916	194
2014	1010	249
2015	1,280	204
2016	937	104
Total	**6,960**	**1,471**

Our review produced a total of 6,960 articles that referred to "social robots." The distribution of the articles by year is shown in Table 1. From these articles, we identified 1,471 studies that examined human responses to specific robots, and that also included an image of the robot. From these studies, our sampling produced a total of 342 specific social robots. This collection of robots reasonably represents the current population of machines that operationally define this category of technology, at least with respect to the robots used in research. All of the robots are shown with small icons in Figure 2.

The entire collection of robots represents significant breadth of design and substantial differences across the robots on qualities known to be influential in social and person perception generally. As a preview to our empirical analysis of different robot impressions and attributes, we mention literature that suggests the most important features of robots that could be coded and analyzed. For example, an eye tracking study found that humans pay most attentions to the head area of the social robots (Dziergwa et al., 2013). For trait perceptions of robots, researchers favor composites of facial and bodily features in single studies. Researchers have compared human-like versus machine-like presentations that use collections of different attributes (Bartneck et al., 2009; Broadbent et al., 2013; DiSalvo et al., 2002; Fiske, 2015; Goetz et al., 2003; Haring et al., 2016; Hegel et al., 2008; Li et al., 2010; Phillips et al., 2017; Walters et al., 2009). In those studies, human-like included eye size, facial features, and human body shapes; machine-like included shapes, text, and movements. Others studies have examined categories of features, for example, those associated with animals (Lakatos et al., 2014; Lohse et al., 2007; Miklósi & Gácsi, 2012; Yanco & Drury, 2004). Isolated attributes are occasionally studied, including examination of overall height, head size, arm length, and skin color, all predictors of believability (Bogdanovych et al., 2016), and chin size, gender, movement, and masculinity, all found to be related to impressions of social robots (Lehmann et al., 2015; Powers & Kiesler, 2006). Fischer et al. (2012) showed that physical embodiment influenced how much a robot is perceived as an interaction partner and showed degrees of freedom in motion affected how users evaluate the suitability of a robot for different tasks.

Four studies have used collections of robots to investigate the link between their physical characteristics and human perceptions.

Figure 2
Images of all Social Robots From 2005 to 2016
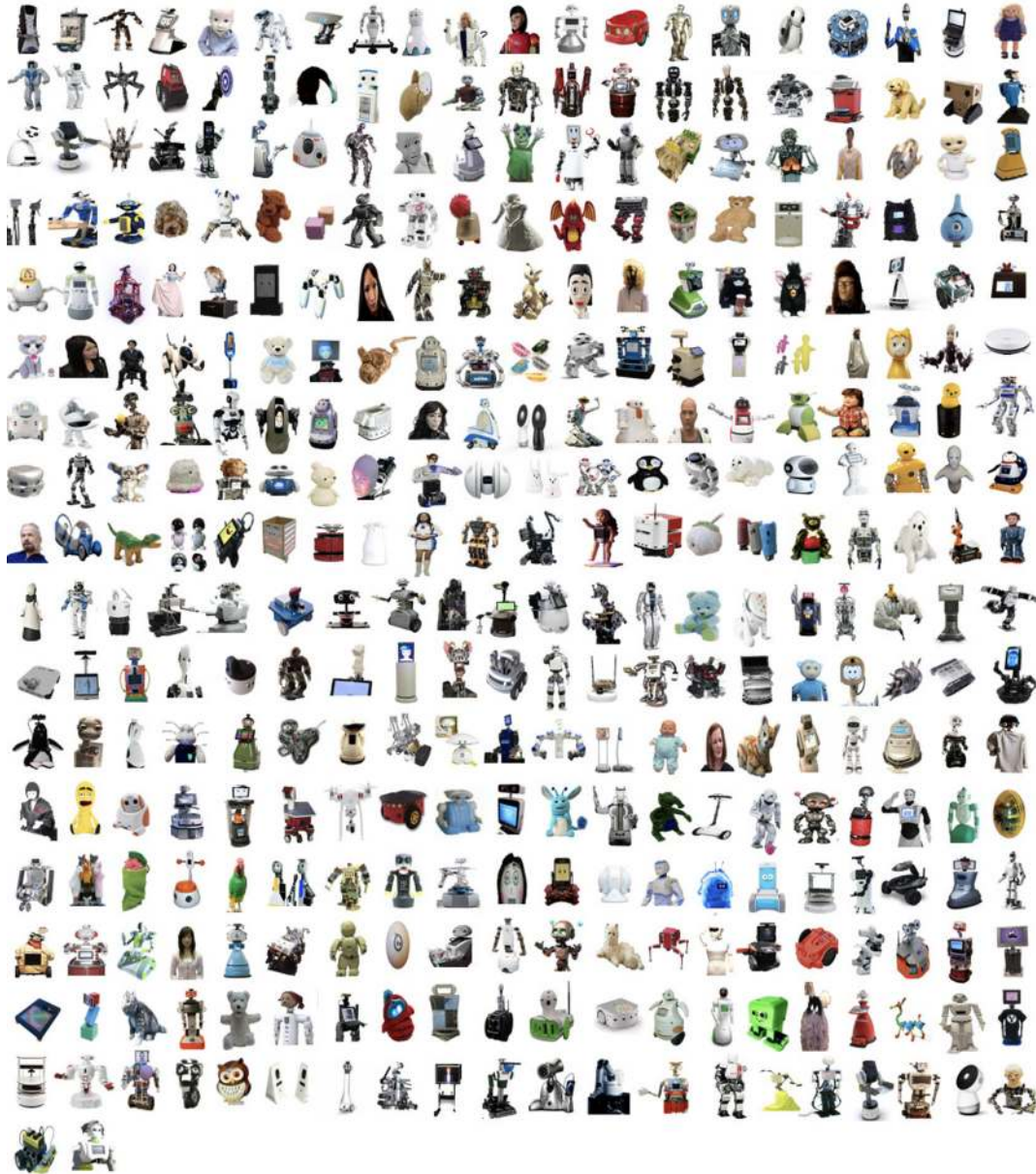
von der Pütten and Krämer (2012) studied 40 images of robots and found that tall and bipedal robots invoked feelings of threat. Disalvo et al. (2002) used images of 48 robots to study robot humanness and showed that the dimensions of the head and the total number of facial features influenced the perception of humanness. Kalegina and colleagues (2018) build a database of rendered robot faces and explored features that included variations in the mouth, nose, eyebrows, colors, eye size, cheeks/blush, and face shape. They surveyed participants and asked them to assess the robots on several measures (machine-like–human-like, unfriendly–friendly, unintelligent–intelligent, untrustworthy–trustworthy, childlike–mature, and masculine–feminine). They found that the faces with

no pupils and no mouth were consistently ranked as unfriendly, machine-like, and unlikable. Robots with pink or cartoon-styled cheeks were consistently ranked as feminine across both studies. Phillips and colleagues (2017) analyzed 155 drawings of robots to understand a priori expectations about robot appearance. They found that people's visualizations of robots have common attributes including human-like motion, human facial and body features, and gendered appearance. In another study, Phillips and colleagues collected and analyzed 200 images of robots with at least one human-like feature and they found four distinct appearance dimensions that characterize anthropomorphic robots: Surface look (eyelashes, head hair, skin, genderedness, nose, eyebrows, and apparel),

body manipulators (hands, arms, torso, fingers, and legs), facial features (face, eyes, head, and mouth), and mechanical locomotion (wheels and treads/tracks).

From this research and our own examination of the robots in our database, we developed a collection of 21 attributes that we applied to the sample of the 342 robots. These features are listed in Table 2 and include descriptive attributes (e.g., has a face or not and mechanics are visible) and subjectively perceived attributes (e.g., masculinity and femininity and age).

Method

We reviewed our entire database of social robots to identify the physical characteristics considered in the design of social robots. We randomly selected 10% of the sample of social robot photos to develop the coding scheme that included the 21 attributes described in Table 2. Thirteen of these attributes are descriptive and are coded by our research team, including whether the robot has a face or not, has vision or not, its skin type is (e.g., furry and metallic), shape (humanoid or not), and capacity for motion (degrees of freedom, locomotion, and bipedal). Eight attributes required perceptual judgments, including perceived gender, masculinity, femininity, developmental category, eye/head ratio, head/body ratio, height, and age. We recruited participants to provide their evaluations on these eight attributes.

A total of 4,415 participants recruited from Amazon's Mechanical Turk service evaluated each robot on the eight attributes. Of the participants, 54.9% were male and 44.0% were female,

0.5% were not binary, and 0.6% prefer not to report their gender. The mean age of the participants was 39.3 ($SD = 11.7$) and 78.7% of the participants were White, followed by Asian (10.12%), African American (8.08%), Latino (5.07%), and others (2.08%).

Each participant was randomly assigned to rate one robot, and each robot received at least 10 ratings. The ratings were averaged to produce a score for each item for each robot (Salganik, 2019). The minimum of 10 evaluations was allowed for stable averaging across the evaluations for each robot. We calculated the mean of the 10 scores from participants on the six continuous variables: eye/head ratio, head/body ratio, height, masculinity score, femininity score, and age. We calculated the mode on two categorical variables: gender and age categories (data available at https://shorturl.at/zACIP). We have also made the study materials, data, and analytic methods publicly available.

Results

We calculated the Pearson correlations among the 21 attributes (see Table 3) and then conducted a principle component analysis to explore the connections among the 21 variables. We used categorical principal components analysis (CATPCA) with Varimax rotation to explore grouping dimensions of the coded social robot physical characteristics. The results explained 34.82% variance. The data were grouped into two dimensions, one machine-like and one human-like. The first factor, machine-like, contained physical characteristics of metal skin, locomotion, size, masculinity, femininity, age, degrees of freedom, fur skin, animal shape, and visibility of mechanics. The second factor, human-like, had items related to face, vision, speech, age group, eye versus body ratio, and head versus body ratio (Table 4).

Discussion

The entire collection of robots represents significant breadth of design and shows substantial differences across the robots on qualities known to be influential in social and person perception generally. Estimated sizes ranged from 0.3 to 6.0 ft; visibility of electronics and mechanical features was obvious for some and hidden for others; materials could be anything from soft fur to human-like plastic to metal; there were caricatures of humans and some with uncanny human similarity; and facial features from puppet-like and humorous to serious all-business expressions.

A main purpose of the literature review was to discover the variance in social robots used in research that should be represented in empirical studies about how robots are perceived. If the sampling of robots is considered as a random factor in research, as we have argued it should be, then this is the variance that needs to be sampled to allow the best generalizations possible for models of robots and social dynamics. The 342 robots found in the literature are a useful collection to bring to new research. Before reporting psychological responses to this collection, it is possible to make several comments about the robot collection per se. First, the literature, almost seven thousand citations, is large and now comprises over 3,000 separate studies. This is largely because of the promise, mentioned in the introduction to almost every article we reviewed, that social robots may help solve important social problems, mostly in the areas of education and healthcare.

Table 2

Features of Social Robots

Features	Coding type	Descriptive statistics
Head		
Eye:head ratio	Perceived	$M = .11, SD = .06$
Head:body ratio	Perceived	$M = .15, SD = .07$
Has face	Descriptive	No: 32.5%, yes: 67.5%
Has vision	Descriptive	No: 30.7%, yes: 69.3%
Skin type and shape		
Plastic	Descriptive	No: 39.8%, yes: 60.2%
Metal	Descriptive	No: 51.8%, yes: 28.2%
Fur	Descriptive	No: 88.6%, yes: 11.4%
Silicone	Descriptive	No: 90.4%, yes: 9.6%
Mechanics visible	Descriptive	No: 59.4%, yes: 40.6%
Animal shape	Descriptive	No: 79.5%, yes: 20.5%
Height	Perceived	$M = 3.02, SD = 1.38$
Communication ability		
Has speech	Descriptive	No: 50.6%, yes: 49.4%
Has digital presentation	Descriptive	No: 73.7%, yes: 26.3%
Motion		
Degrees of freedom	Descriptive	$M = 4.35, SD = 4.67$
Locomotion	Descriptive	No: 34.8%, yes: 58.8%, cannot tell: 6.4%
Bipedal	Descriptive	No: 76.3%, yes: 23.7%
Gender		
Gender displayed	Perceived	Male: 42%, female: 12.2%, not displayed: 45.8%
Masculinity	Perceived	$M = 53.62. SD = 19.12$
Femininity	Perceived	$M = 33.75, SD = 20.02$
Age		
Developmental category	Perceived	Child: 42.7%, adult: 32.5%, senior: 0.9%
Age	Perceived	$M = 21.65, SD = 11.43$

Table 3
Correlations Among 21 Attributes

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
1. Has a face																				
2. Bipedal	.14*																			
3. Metal skin	-.19**	-.10																		
4. Vision	.69**	.12*	-.23**																	
5. Plastic skin	-.07	-.01	-.01	-.02																
6. Speech	.50**	.22**	-.19**	.54**	-.07															
7. Locomotion	-.05	.12*	.06	-.08	.10	-.06														
8. Age Range	-.46**	-.21**	.09	-.49**	-.19**	-.31**	.06													
9. Height	.09	.21**	.21**	.12*	-.01	.14*	.26**	-.09												
10. Ratio of head and body size	.51**	.08	-.23**	.51**	-.03	.38**	-.25**	-.49**	-.17**											
11. Gender display	-.01	-.03	.01	.00	-.03	-.06	.01	.10	.03	-.09										
12. Masculinity	-.20**	.13*	.27**	-.23	.03	-.11*	.11*	.15*	.16**	-.22**	.04									
13. Femininity	.19**	-.05	-.25**	-.24**	-.04	.16**	-.13*	-.13*	-.05	-.22**	-.04	-.89**								
14. Age score	-.10	.10	.20**	-.15**	-.08	.02	.13*	.25**	.51**	-.30**	.05	.24**	-.08							
15. Degrees of freedom	.12**	.39**	.16**	.08	.09	.07	.23**	-.15*	.25**	-.05	.02	.26**	-.22**	.08						
16. Ratio of eyes and head size	.55**	.10	-.05	.68**	.03	.40**	-.01	-.49**	.13*	.49**	-.06	-.09	.09	-.13**	.10					
17. Fur skin	.17**	-.07	-.33**	.18**	-.37**	.12*	-.19**	.07	-.30**	.29**	-.06	-.12*	.09	-.23**	-.16**	.04				
18. Rubber silicone skin	.06	.19**	-.24**	.11*	-.18**	.09	-.05	0	.18*	.01	.08	-.05	.09	.10	-.01	-.02	-.12*			
19. Animal shape	.17**	-.10	-.27**	.15**	-.23**	.14*	-.19**	.07	-.39**	.30**	-.09	-.06	.04	-.24**	-.08	.08	.62**	-.09		
20. Digital presentation	.02	-.08	.02	.01	-.02	.02	.12*	.05	.06	-.08	.06	.02	-.04	.02	-.09	.01	.02	-.06	.03	
21. Visibility of mechanics	-.15**	-.01	.42**	-.10	.02	-.08	.03	.10	.23**	-.24**	.06	.23**	-.21**	.21**	.22**	-.04	-.28**	-.07	-.20**	.07

Table 4
Factor Loadings of 21 Features of Robots

	Component 1 (machine-like)	Component 2 (human-like)
Has a face (1: no; 2: yes)	−0.223	**0.771**
Bipedal (1: no; 2: yes)	0.260	0.379
Metal skin (1: no; 2: yes)	**0.533**	−0.174
Vision (1: no; 2: yes)	−0.239	**0.822**
Plastic skin (1: no; 2: yes)	0.194	0.036
Speech (1: no; 2: yes)	−0.151	**0.662**
Locomotion (1: no; 2: yes)	**0.509**	−0.107
Age range (1: child; 2: adult; 3: senior)	−0.196	**−0.721**
How many feet	**0.640**	0.353
Ratio of head and body size	−0.507	**0.582**
Gender display (1: male; 2: female; 3: not displayed)	0.103	−0.073
Masculinity score (1–100)	**0.537**	−0.158
Femininity score (1–100)	**−0.479**	0.189
Age score (1–100)	**0.533**	−0.10
How many degrees of freedom	**0.519**	0.302
Ratio of eyes and head size	−0.082	**0.742**
Fur skin (1: no; 2: yes)	**−0.655**	−0.007
Rubber/silicone skin (1: no; 2: yes)	0.026	0.181
Animal shape (1: no; 2: yes)	**−0.638**	−0.011
Has digital presentation (1: no; 2: yes)	0.037	−0.034
Visibility of mechanics (1: no; 2: yes)	**0.527**	−0.057

Second, only a minority (21%) of the published research we examined measured human responses to robots. Much of the research is devoted to proposals for different uses (e.g., this is how a social robot might assist with elder care), and specifications for different mechanical or computing features of the robots. It is understandable that much of the literature should address questions about how to build the machines, and where and when they might be useful. Progress on the grand challenge of building machines that simulate human social interaction, however, will require more empirical work that measures human responses and participation in the interactions.

Third, the group of robots used in research with humans is impressive in breadth. For us and for colleagues who have seen Figure 2, social robots are considerably more varied than most think, even if you follow the literature. One cannot look at Figure 2 without concluding that radically different machines make up a technology category that many are comfortable defining with exact boundaries, in spite of the variance within the category. Also, there is no single robot or even small group of them that dominate the literature. Although some robots have been used in multiple studies (e.g., Paro, AIBO, Nao, Robovie, and iCat), the number of citations using the same robot was low (with the exception of Nao which was used in 225 papers or 15% of the papers with robots). The top five most studied robots in addition to Nao are Paro (73 papers, 5%), Robovie (71 papers, 5%), AIBO (66 papers, 5%), and iCat (59 papers, 4.0%).

The large variance in robots is both a problem and an opportunity. The opportunity is that the variance allows for better definitions of the attributes that might differentiate robots, and discovery of those attributes should be useful in discussions about designing new ones. The dimensions that might emerge from new research will be more accurate because they will more completely represent robot possibilities. The problem with the large variance, however, is that it signals concerns about stimulus sampling should be paramount. Research about a single robot, an attribute of many of the studies we examined, severely limits the generalization of research about social robots to other robots not included in the research.

Study 2: Warmth and Competence

What are the most important descriptions of the 342 robots? Two questions help define a path to an answer. The first is whether there is something special about mechanical robots that require an understanding of how humans might perceive them as a special category of objects. Robots are certainly special for the people who build and design them, requiring new ideas about materials, electronics, mechanics, and software. But are they unique as social actors in social interactions? Is there something about their exaggerated attributes, their obvious machine characteristics, or the contexts in which they operate that requires a separate psychology of robots?

We think that the answer to this question, based on a considerable literature, is no. With respect to features of social interaction, research shows that human responses to technology, including robots, are fundamentally social and natural, just like human responses to other people (Moon & Kim, 2001; Reeves & Nass, 1996). Technology is sufficiently human-like that people use the same perceptual strategies and biases that exist for negotiating real human interactions. This has been found across a broad range of social characteristics; for example, with respect to the personalities of television characters (Hoffner & Buchanan, 2005) (they are the same as those for real people); personality of computer interfaces (Nass & Lee, 2001) (introverted and extroverted interfaces cause the same impressions as for real people); politeness rules that people use to evaluate computers and other technology (Nass, 2004) (people are polite to computers in the same ways there are other people); and similar results for several other social evaluations, including reciprocity (Katagiri et al., 2001), gender stereotypes (Lee et al., 2000), specialization of expertise (Koh & Sundar, 2010), and team membership (Nass et al., 1996). One recent study even found that physiological responses to touching robots produced the same differences in arousal levels depending where on a mechanical body people were asked to touch (Li et al., 2017).

If the psychology of human perception applies to robots, then what are the aspects of human social perception that might determine what people think about *any* social actor, human, or robot? Previous research has found several attributes important for robots, including anthropomorphism (Kamide et al., 2013), familiarity (Baddoura & Venture, 2013; Baddoura et al., 2012), perceived intelligence (Haring et al., 2016; Hegel et al., 2008), sociability, likeability (Li et al., 2010), and trustworthiness (Mathur & Reichling, 2016). Researchers have also developed measures to capture how people perceive robots. Kamide et al. (2014) developed the PERNOD (PERception to humaNOiD) scale that measures the attribute of humanoid. They found five basic dimensions for perceiving humanoid robots: familiarity, utility, motion, controllability, and toughness. Similarly, Bartneck et al. (2009) identified five basic attributes that professionals and researchers use to evaluate robots: anthropomorphism, animacy, likeability, perceived intelligence, and perceived safety. Gray et al. (2007) proposed two dimensions of perceptions of robots: agency and experience. Agency referred to

the capacity to act and experience referred to ability to experience feelings and emotions.

Across these robot studies, there seem to be two attribute clusters: one reflecting feelings of familiarity, warmth and friendliness, and another representing perceived competence and intelligence. Both clusters compare well to attributes that humans use to form impressions of other humans. The SCM argues that, like all perception, social perception depends on evolutionary pressures, and most importantly the pressure to determine quickly whether another social actor is friend or foe, and whether that social actor has the ability to act on friendly or aggressive intentions (Fiske et al., 2007). Decades of research has verified that these dimensions, named warmth (and also trustworthiness, sincerity, friendliness, helpfulness, or morality) and competence (and also intelligence, skill, creativity, and efficacy), represent the quick evaluations that people make of individuals from presidential candidates, to well-known people, to meetings with strangers, and that these two dimensions appear stable across cultures and over time (Fiske, 2018).

It is important to note that these evaluations are not thought to be opinions about people based on extended interactions or considerable thought and contemplation. The judgments of warmth and competence are unconscious, automatic, and reflexive and are the basis of stereotypes we hold for various categories of people (Eyssel & Hegel, 2012). They are made because of the constraints imposed by evolutionary pressures to survive, yet they are applicable to contemporary social interactions even though the original situations that favored their evolution are no longer as important as they were in our distant past. Importantly for application to robots, the trait judgments that people make in response to faces occur in an instant, as quickly as 33–38 ms after exposure to a face, even if the face is chosen to be as emotionally neutral as possible (Bar et al., 2006). These are primitive, quick, and consequential judgments.

The SCM builds on these two core evaluation dimensions to predict the content of stereotypes that people have of others (Fiske et al., 2002). The model describes four quadrants based on combinations of the dimensions. For example, people perceived as high in warmth but low in competence evoke paternalistic stereotypes (e.g., elderly people) while people low in warmth but high in competence evoke envious stereotypes (e.g., rich people). People judged low in both warmth and competence evoke contemptuous stereotypes (e.g., welfare recipients) while those judged high in both warmth and competence evoke admiration (e.g., ingroup and allies) (Fiske et al., 2002).

Our question in Study 2 was whether people use these same two dimensions to differentiate a broad range of social robots. The same evaluations have been used occasionally in the past research to comment on particular attributes of robots (Bergmann et al., 2012; Eyssel & Hegel, 2012), but no research has looked at their applicability across a large sample of robots. Our first test of human responses, based on evaluations from 3,920 people evaluating all of the 342 robots, used standard warmth and competence evaluative scales from human studies (Fiske et al., 2002).

Knowing that impressions are quickly and reliably defined by warmth and competence can alert a robot designer to important parameters that will determine the success of their machines. Knowledge of the parameters, however, does not determine exactly how to build the machines. In Study 2, we try to answer that question: What does a warm or competent robot look like?

Method

Our goal was to obtain a large human sample for the evaluations of the robots to allow averaging across the evaluations for each robot. We recruited $N = 3,920$ participants from Amazon's Mechanical Turk. This study was approved by Stanford Institutional Review Board. Participants were randomly assigned to the robot images resulting in a minimum of 10 evaluations for each robot (some robots received more than 10). Each participant was presented with a single image of one of the 342 social robots. Participants then rated their own perceptions of the social robot on the nine-item warmth and competence scale (Fiske et al., 2002). The items included tolerant, warm, good natured, sincere, competent, confident, independent, competitive, and intelligent, with the participant rating the social robot on each dimension on a five-point Likert scale, from "not at all" to "extremely." Following our method in Study 1, the ratings were averaged to produce a score for each item for each robot. We have made the study materials, the data, and analytic methods publicly available. We were not able to collect demographic information about the participants who participated via Mechanical Turk, which is a limitation of the current study.

We then used the ratings of the robot attributes to explore how the ratings predicted perceived warmth and competence. Specifically, we used the ratings of perceived warmth and perceived competence of social robots as the dependent variables and used the judgments about attributes from Study 1 as predictors (data available at https://shorturl.at/sFJO3).

Results

The standard analysis strategy for examining the dimensionality of a set of social actors is to factor analyze several evaluations, paying attention to whether the traits cluster within the larger categories of warmth and competence, and to how much of the total variance across the evaluations can be explained by these two dimensions as opposed to others. Although these warmth and competence items have been used in numerous studies applied to the perceptions of humans (Fiske, 2018), the application of these items to the perceptions of a large sample of social robots is novel. We therefore conducted an exploratory factor analysis (EFA) with principle axis factoring with oblique rotation to examine whether the dimensions of warmth and competence emerged. The analysis produced a two-factor solution. The first factor was comprised of the four warmth items (*tolerant, warm, good natured,* and *sincere*) ($M = 3.38$, $SD = .39$, $\alpha = 0.89$) that explained 31% of the variance. The second factor was comprised of the five competence items (*competent, confident, independent, competitive*, and *intelligent*) ($M = 3.46$, $SD = .30$, $\alpha = 0.82$) that explained 30% of the variance. The factor loadings are described in Table 5.

This two-factor structure for perception of social robots is consistent with the warmth-competence factor structure from human studies described in the SCM (Fiske, 2015, 2018) both in the relative ordering of the components of each dimension, and in the total relative variance explained by the two dimensions.

To answer the second research question, which is how perceived warmth and perceived competence can be explained by the attributes of the robots, we first used the machine-like factor score and the human-like factor score generated from Study 1 to predict perceived warmth and perceived competence. We conducted a

Table 5

Item Loadings for Robot Perceptions on the Warmth-Competence Scale

Item	Component 1 (warmth)	Component 2 (competence)	Mean (SD)
Competent	−.11	**.85**	3.65 (.38)
Tolerant	**.60**	.27	3.57 (.37)
Confident	.23	**.69**	3.58 (.38)
Warm	**.92**	−.13	3.11 (.55)
Independent	.09	**.64**	3.41 (.36)
Good natured	**.94**	−.01	3.50 (.47)
Competitive	−.12	**.49**	2.98 (.40)
Sincere	**.78**	.17	3.35 (.40)
Intelligent	.00	**.81**	3.68 (.41)

linear regression and found that for perceived warmth, human-like positively predicted perceived warmth ($\beta = 0.33$, $SE = 0.02$, $p < .001$), while machine-like negatively predicted it ($\beta = −0.30$, $SE = 0.019$, $p < .001$). The regression was significant $F(2, 339) = 43.32$, $p < .001$, with the two factors accounting for 19.9% of the variance in perceived warmth. Tests to see if the data met the assumption of collinearity indicated that multicollinearity was not a concern (human-like, tolerance = 1.0 and Varience Inflation Factor (V) = 1.0; machine-like, tolerance = 1.0 and Varience Inflation Factor = 1.0).

We conducted a second linear regression with perceived competence as the dependent variable. Machine-like was significantly associated with perceived competence ($\beta = 0.42$, $SE = 0.015$, $p < .001$), while human-like was not ($\beta = 0.033$, $SE = 0.015$, $p = .51$). The regression was significant $F(2, 339) = 35.53$, $p < .001$ and these two factors explained 16.8% of the variance in perceived competence. Tests to see if the data met the assumption of collinearity indicated that multicollinearity was not a concern (human-like, tolerance = 1.0 and VIF = 1.0; machine-like, tolerance = 1.0 and VIF = 1.0).

We then used the attributes to predict perceived warmth and competence. Two separate linear multiple regressions were conducted with perceived warmth and competence as the dependent variables. We used multiple dummy codes for the gender displayed, developmental categories, and locomotion.

A robot was perceived with higher warmth if its mechanics were invisible ($\beta = −0.13$, $SE = 0.05$, and $p < .05$), it had more degrees of freedom for movement ($\beta = 0.12$, $SE = 0.01$, $p < .05$), younger age ($\beta = −0.14$, $SE = 0.01$, $p < .05$), and displayed as a child ($\beta = −0.14$, $SE = 0.01$, $p < .05$). The regression was significant $F(25, 316) = 5.90$, $p < .001$ with the four factors explaining 26.4% of the variance in perceived warmth. The standardized beta coefficients are presented in Table 6.

Table 6

Robot Features Predicting Warmth and Competence Regression

Features	Type	Warmth β	Warmth SE	Competence β	Competence SE	Tolerance	VIF
Head							
Eye:head ratio	Continuous	0.10	0.45	0.00	0.36	0.44	2.26
Head:body ratio	Continuous	0.04	0.38	−0.01	0.31	0.46	2.16
Has face	Categorical	0.04	0.06	−0.12	0.05	0.44	2.29
Has vision	Categorical	0.14	0.07	0.10	0.06	0.32	3.10
Skin type and shape							
Plastic	Categorical	0.01	0.05	−0.04	0.04	0.67	1.50
Metal	Categorical	−0.04	0.05	0.05	0.04	0.56	1.78
Fur	Categorical	0.05	0.09	−0.24**	0.07	0.45	2.22
Silicone	Categorical	−0.05	0.07	−0.01	0.06	0.70	1.43
Mechanics visible	Categorical	−0.13*	0.04	−0.08	0.04	0.72	1.39
Animal shape	Categorical	−0.01	0.06	−0.13	0.05	0.52	1.94
Height	Continuous	−0.03	0.02	0.01	0.02	0.44	2.27
Communication ability							
Has speech	Categorical	−0.03	0.05	−0.02	0.04	0.62	1.61
Has digital presentation	Categorical	0.09	0.04	0.07	0.03	0.93	1.07
Motion							
Degrees of freedom	Continuous	0.12*	0.01	0.24*	0.00	0.66	1.51
Locomotion no vs. yes	Categorical	0.11	0.08	0.00	0.07	0.21	4.83
Locomotion no vs. not sure	Categorical	0.16	0.08	0.07	0.07	0.20	4.94
Bipedal	Categorical	0.02	0.05	0.01	0.04	0.71	1.40
Gender							
Gender displayed: not displayed vs. male	Categorical	0.04	0.04	−0.01	0.03	0.85	1.18
Gender displayed: not displayed vs. female	Categorical	−0.08	0.06	0.05	0.05	0.85	1.18
Masculinity	Continuous	−0.15	0.00	−0.08	0.00	0.17	5.80
Femininity	Continuous	0.00	0.00	−0.09	0.00	0.18	5.56
Age							
Developmental category: child vs. adult	Categorical	−0.11	0.06	0.03	0.05	0.45	2.24
Developmental category: child vs. senior	Categorical	0.02	0.22	0.05	0.18	0.75	1.33
Developmental category: child vs. not applicable	Categorical	−0.14*	0.06	0.01	0.05	0.44	2.26
Age	Continuous	−0.14*	0.00	0.04	0.00	0.46	2.19
Adjusted R^2		**0.24**		**0.20**			

Figure 3

Sample of 100 Social Robots in Warmth-Competence Space Taken From the Population of 342 Robots in the Published Literature From 2005 to 2016 Where Each Robot Is Placed Along the Dimensions of Competence and Warmth
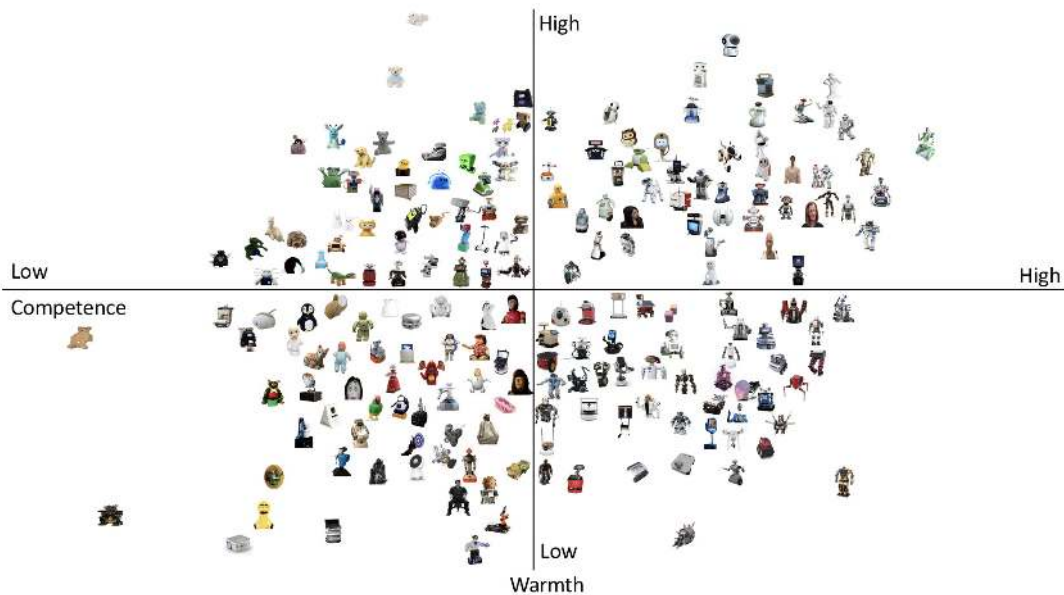
A robot was perceived as a more competent the more degrees of freedom for movement ($\beta = 0.24$, $SE = 0.00$, $p < .01$) and if it did not have fur skin ($\beta = -0.24$, $SE = 0.07$, $p < .01$). The regression was also significant $F(25, 316) = 4.06$, $p < .001$. The two factors explained 18.3% of the variance in perceived competence. We reviewed the tolerance and variance inflation scores to check multicollinearity and found that masculinity and femininity had VIF scores of 5.80 and 5.57, respectively. The rest of the variables did not exhibit multicollinearity. Figure 3 shows how a sample of 100 robots from our larger collection is arrayed in the two-dimensional space of warmth and competence.

Discussion

These results show that perceptions across a large sample of social robots are like those for real people. The evolutionary pressures that shaped evaluations of warmth and competence, developed thousands of years before friends and foes could be anything other than humans or animals, are nonetheless applicable to the evaluation of 21st century technology. Social robots engage similar social perception and impression formation processes as real people. This is supported in spite of obvious differences in their appearance, and especially in the appearance of extreme attributes that often are their most desirable and purposive qualities (e.g., wheels for legs, digital displays in place of a face, metallic surfaces with visible internal wiring, and unusual animal-like forms).

Although the primacy of these evaluations, applied both to humans as well as nonhuman forms, may seem uncontroversial, we note that quite often robots are given special psychological status. For example, many studies, and especially popular conversations about robots, discuss how robots may form special relationships over extended time periods (Gockley et al., 2005; Leite et al., 2013) or how repetition in behaviors may cause boredom or evaluations of unnaturalness (Baxter et al., 2011). Although these more elaborate and slower responses are possible, and may be important for evaluations about how or whether relationships with a robot may change and endure over time, it is the faster automatic responses that will set the basic parameters for how people respond to the machines. Warmth and competence evaluations should determine whether people will walk toward a robot or back away, whether friendly or aggressive intentions will be assumed, or whether the intensity of any of those evaluations should be increased because a robot appears to have the means to follow through on whatever behaviors seem most possible.

One limitation of this study is that impressions of the 342 robots were based on looking at pictures. How might results change if people were responding to the actual machines, and watching them move and interact? There is no doubt that much more information about the robots would be available during live interactions, and where possible, research about social responses should include as much richness in the interaction as possible. It is also worth noting, however, that the primitive interactions considered here happen quickly (Willis & Todorov, 2006). When people are shown pictures of human faces, the same two dimensions of warmth (trustworthiness) and competence (dominance) are found after extremely brief exposures (Todorov et al., 2008). Further, impressions that are given within 100 ms are highly correlated with evaluations made in the absence of time constraints (Leite et al., 2013). So this is a fast process that is influential from the very beginning of an interaction.

We note three other conclusions. First, we can say that it is possible to successfully predict warmth and competence responses to robots using a mix of attributes similar to human evaluation and ones that are unique to robots. The level of prediction is sizable, suggesting that these attributes can be used as design guidelines with some confidence. Second, for evaluation of warmth, age is a primary predictor. When people perceive a robot as young they are more likely to perceive the robot as warm. In Figure 3, Paro and Pepper are good examples. A second predictor of warmth was mechanical visibility. Warm robots do not have as much visible machinery like wiring, servos, and gears. In Figure 3, good examples of robots of each extreme are Paro and Qin for warm, and Youbot and Mentorbot for cold.

Third, a robot's perceived mobility, in particular its perceived degrees of freedom, predicts both perceived warmth and competence. This was the only feature to predict both warmth and competence. Although further research is required to understand why this might be, it is likely that these perceptions rely on the ability to carry out being either warm or competent. People naturally made evaluations of actionability, a judgment that is both primitive and connected to more obvious subjective judgments about a robot's warmth or competence. Recall, for example, that primitive judgments were about the perceived ability of people we encounter to act regardless of whether the actions would be perceived positively or negatively. We found a comparable judgment for robots.

An important caveat about these judgments about robots is that they are made quickly, within milliseconds. Consequently, a robot that *looks* like it could or could not move to help a human may be an important precursor to more elaborate evaluations that may develop over time—ones that would be determined also by whether or not a robot could actually follow through on initial perceptions. But robots that do not look warm or competent, even if they might perform good naturedly or competently over time, may not get a chance to offer any proof if rejected after first impressions.

Finally, competence is also defined in the social robot population as being nonanimal, at least enough so that fur is not apparent. Contrary to stereotype, the presence of fur does not make a robot more warm.

Stimulus Sampling

The collection of robots shown in Figure 2 highlights the large diversity of social technology in this category and makes salient the issue of stimulus sampling for advancing the field of social robots. The practicality of sampling multiple robots for a single study, however, is obviously difficult. Although psychologists can sample relatively easily from some stimulus categories (e.g., emotional faces and categories of text), it is hard to include multiple robots in experiments because each machine is expensive to produce or purchase, and because different robots, each with their own software, would need to be programmed to do similar tasks. There are, however, methods to accomplish improved sampling with social robots. One possibility is to increase the use of between-study meta-analyses that allow researchers to compare how different robots perform on a given task by comparing them across (rather than within) studies. Another approach is to standardize robots across studies. It may be useful, for example, to build common robots for use in a particular types of research (e.g., teaching young students and creating assistive social robots for a specific population)

(Baxter et al., 2017; Ueyama, 2015). This kind of standardization would make comparison across studies much easier, as would develop a more standardized set of tasks and dependent variables. This approach has led to significant advances in other fields with high diversity with stimulus samples, such as the advance of natural language processing with standardized tasks.

The Stereotyping of Social Robots

In social psychology, the SCM describes how the two dimensions of warmth and competence produce stereotypes about humans that drive emotional and behavioral responses (Fiske, 2015, 2018; Fiske et al., 2002). We looked to see if similar stereotyping might also have occurred with the robots. The SCM lays out the two-dimensional warmth-competence space into four quadrants, with those in the high warmth, high competence quadrant (the upper right in Figure 4) representing stereotypes from the in-group for that specific culture. In the U.S. studies, for example, in-group stereotypes include Americans, Christians, housewives, and the middle class. Figure 4 shows six social robots at the extremes of the high warmth and competence in-group quadrant. They are strikingly similar, and exclusively human-like (no animals and no toys), and they each have four limbs, a head proportionate to human body size, and they appear dominant and even athletic. These may be, at least for our sample of evaluators, the in-group stereotype of social robots.

The other three quadrants in the SCM space describe different out-groups, based on fundamental intergroup biases from social psychology (Fiske, 2015, 2018). For human stereotypes, the quadrant lowest on both dimensions includes homeless people and immigrants. The social robots in that group (line 4 in Figure 4) are some of the least action-oriented of the larger group (e.g., a sad yellow gumby-style figure, a flattened bear), and all appear without noticeable emotion.

The other two quadrants represent more ambivalent stereotypes. Low warmth and high competence people, in research done in the United States, stereotypically include groups like rich people and Asians, and they are viewed with envy and jealousy, and perceived to have prized abilities but suspect intentions. Figure 4 shows six social robots in that quadrant. They are much less human-like, have exaggerated mechanical features (e.g., antennae and spidery legs), sometimes no identifiable body, and wheels instead of legs. These may be attributes that suggest competence, but they also may signal ambiguous intentions. In contrast, high warmth and low competence robots are shown in the third line of Figure 4. Human stereotypes in that quadrant include the elderly and disabled. The corresponding attributes for robots are plush animal bodies (e.g., dogs, bears, and a llama), and toy-like representations of objects.

Importantly, the SCM describes how groups represented in each quadrant evoke different emotional reactions. People high on both warm and competence dimensions (e.g., in-group allies) evoke admiration, while people low in both dimensions (e.g., poor and homeless) evoke disgust. People in the quadrant of low warmth but high competence (e.g., rich people and professionals) evoke envy, while people in the high warmth but low competence quadrant (e.g., older and disabled) evoke pity. Given how closely the evaluations of robots tracked the evaluations people, the SCM suggests that robots in the different quadrants may evoke classical stereotypes,

Figure 4

Sample of Six Social Robots in Each Quadrant of Stereotype Content Model (SCM) Space

Low warmth/low competence   Low warmth/high competence   High warmth/low competence   High warmth/high competence
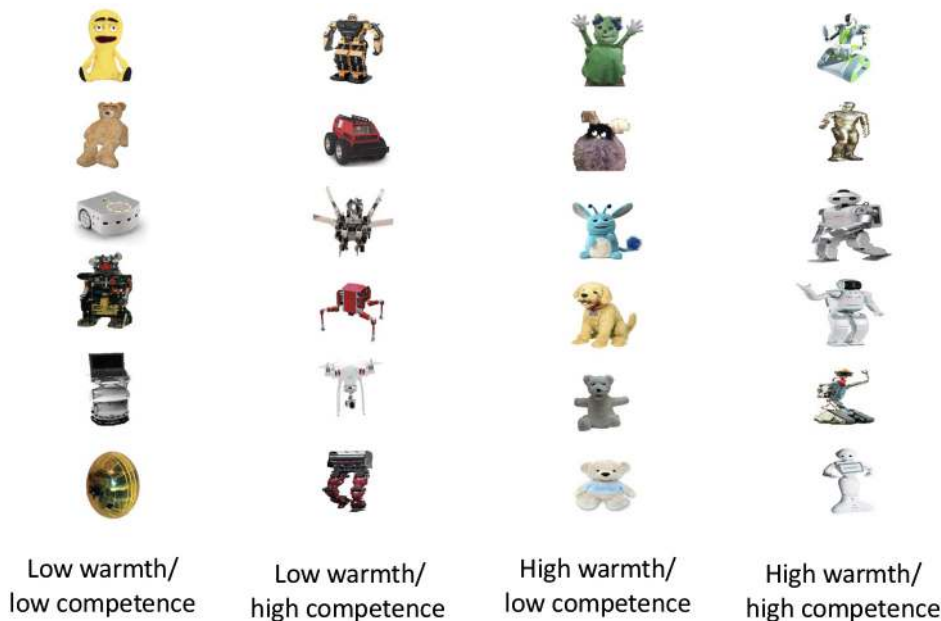
regardless of whether the robot designers had these stereotypes in mind or not.

Finally, the SCM also argues that there are important behavioral implications with respect to each quadrant. High warmth is associated with helping and protecting while low warmth is associated with attacking and fighting. High competence is associated with affiliation while low competence is associated with neglect. Combinations of the two dimensions could create behavioral responses to the social robots that could dramatically change their value in the primary contexts in which they are considered.

Warmth and Competence for Design

Now that there is evidence that warmth and competence are important evaluations of robots, and we know some of the attributes that encourage those judgments, we can think about implications for the design of robots. Based on the extensive literature about warmth and competence judgments about real people (Fiske et al., 2002), what does this mean for the design of social robots?

First, according to the impression literature, warmth judgments are primary and quick (Fiske, 2018). That trait is judged before competence, presumably because the determination of another social actor's friend or foe intent is more important than the judgment about whether that actor will be able to deliver on the intentions. This suggests that if signaling warmth is critical, those attributes most predictive of warmth in robots (e.g., surface materials and perceived age) should be emphasized unambiguously. Judgments about warmth may predict success of an entire relationship. Examples of good models from Figure 3 include Paro, Qin, and Autom.

Second, warmth and competence judgments drive the perception of stereotypes and associated emotional and behavioral reactions described by the SCM. Regardless of whether a designer of robots

tries to use or change these responses to stereotypes, it is clear that the worlds of social robots and human social actors may not be very different. Our data suggest that human partners will perceive a robot's warmth and competence to determine the robot's social standing. Researchers and designers must explicitly reject drawing on problematic stereotypes associated with racism and sexism in the design of social robots. These stereotypes will guide how people will interact with the robot, with crucial implications for the social dynamics of the interaction.

Conclusion

What makes a social robot different or similar to another robot? Our collection of 342 social robots reveals that social robots are remarkably diverse in appearance. Nonetheless, people's perceptions of this large sample of robots conformed to the same primary and rapid evaluations we make of other humans, suggesting that people differentiate robots along the dimensions of warmth and competence, with specific design attributes determining where in that two-dimensional space a robot will be located. In her overview of research examining person perception, Fiske (2018) details how the primacy of warmth and competence spans cultures and has endured over time. Our research indicates that perceptions of warmth and competence extend to our understanding of robots.

References

Baddoura, R., & Venture, G. (2013). Social vs. useful HRI: Experiencing the familiar, perceiving the robot as a sociable partner and responding to its actions. *International Journal of Social Robotics*, *5*(4), 529–547. https://doi.org/10.1007/s12369-013-0207-x

Baddoura, R., Venture, G., & Matsukata, R. (2012). The familiar as a key-concept in regulating the social and affective dimensions of HRI. *2012*

12th IEEE-RAS International Conference On Humanoid Robots (Humanoids) (pp. 234–241). IEEE.

Bar, M., Neta, M., & Linz, H. (2006). Very first impressions. *Emotion*, *6*(2), 269–278. https://doi.org/10.1037/1528-3542.6.2.269

Bartneck, C., Kulić, D., Croft, E., & Zoghbi, S. (2009). Measurement instruments for the anthropomorphism, animacy, likeability, perceived intelligence, and perceived safety of robots. *International Journal of Social Robotics*, *1*(1), 71–81. https://doi.org/10.1007/s12369-008-0001-3

Baxter, P., Ashurst, E., Read, R., Kennedy, J., & Belpaeme, T. (2017). Robot education peers in a situated primary school study: Personalisation promotes child learning. *PLOS ONE*, *12*(5), Article e0178126. https://doi.org/10.1371/journal.pone.0178126

Baxter, P., Belpaeme, T., Canamero, L., Cosi, P., Demiris, Y., Enescu, V., Hiolle, A., Kruijff-Korbayova, I., Looije, R., Nalin, M., Neerincx, M., Sommavilla, G., Tesser, F., & Wood, R., (2011). *Long-term human–robot interaction with young users* [Conference session]. IEEE/ACM Human–Robot Interaction 2011 Conference (Robots with Children Workshop), Lausanne, Switzerland.

Bergmann, K., Eyssel, F., & Kopp, S. (2012). A second chance to make a first impression? How appearance and nonverbal behavior affect perceived warmth and competence of virtual agents over time. *International Conference on Intelligent Virtual Agents* (pp. 126–138). Springer.

Bogdanovych, A., Trescak, T., & Simoff, S. (2016). What makes virtual agents believable? *Connection Science*, *28*(1), 83–108. https://doi.org/10.1080/09540091.2015.1130021

Broadbent, E., Kumar, V., Li, X., Sollers, J. III, Stafford, R. Q., MacDonald, B. A., & Wegner, D. M. (2013). Robots with display screens: A robot with a more human-like face display is perceived to have more mind and a better personality. *PLOS ONE*, *8*(8), Article e72589. https://doi.org/10.1371/journal.pone.0072589

Bruce, A., Nourbakhsh, I., & Simmons, R. (2002). The role of expressiveness and attention in human–robot interaction. *Proceedings of the 2002 IEEE International Conference on Robotics and Automation (Cat. No.02CH37292)* (Vol. *4*, pp. 4138–4142). https://doi.org/10.1109/ROBOT.2002.1014396

Cabibihan, J.-J., Javed, H., Ang, M. Jr., & Aljunied, S. M. (2013). Why robots? A survey on the roles and benefits of social robots in the therapy of children with autism. *International Journal of Social Robotics*, *5*(4), 593–618. https://doi.org/10.1007/s12369-013-0202-2

DiSalvo, C. F., Gemperle, F., Forlizzi, J., & Kiesler, S. (2002). All robots are not created equal: The design and perception of humanoid robot heads. *Proceedings of the 4th Conference on Designing Interactive Systems: Processes, Practices, Methods, and Techniques* (pp. 321–326). ACM.

Dziergwa, M., Frontkiewicz, M., Kaczmarek, P., Kedzierski, J., & Zagdańska, M. (2013). Study of a social robot's appearance using interviews and a mobile eye-tracking device. *International Conference on Social Robotics* (pp. 170–179). Springer.

Eyssel, F., & Hegel, F. (2012). (S)he's got the look: Gender stereotyping of robots. *Journal of Applied Social Psychology*, *42*(9), 2213–2230. https://doi.org/10.1111/j.1559-1816.2012.00937.x

Fischer, K., Lohan, K. S., & Foth, K. (2012). Levels of embodiment: Linguistic analyses of factors influencing HRI. *Proceedings of the Seventh Annual ACM/IEEE International Conference on Human–Robot Interaction* (pp. 463–470). ACM.

Fiske, S. T. (2015). Intergroup biases: A focus on stereotype content. *Current Opinion in Behavioral Sciences*, *3*, 45–50. https://doi.org/10.1016/j.cobeha.2015.01.010

Fiske, S. T. (2018). Stereotype content: Warmth and competence endure. *Current Directions in Psychological Science*, *27*(2), 67–73. https://doi.org/10.1177/0963721417738825

Fiske, S. T., Cuddy, A. J., & Glick, P. (2007). Universal dimensions of social cognition: Warmth and competence. *Trends in Cognitive Sciences*, *11*(2), 77–83. https://doi.org/10.1016/j.tics.2006.11.005

Fiske, S. T., Cuddy, A. J., Glick, P., & Xu, J. (2002). A model of (often mixed) stereotype content: Competence and warmth respectively follow from perceived status and competition. *Journal of Personality and Social Psychology*, *82*(6), 878–902. https://doi.org/10.1037/0022-3514.82.6.878

Gockley, R., Bruce, A., Forlizzi, J., Michalowski, M., Mundell, A., Rosenthal, S., Sellner, B., Simmons, R., Snipes, K., Schultz, A. C., & Wang, J. (2005). Designing robots for long-term social interaction. *In Intelligent Robots and Systems, 2005. (IROS 2005). 2005 IEEE/RSJ International Conference, 2005* (pp. 1338–1343). IEEE.

Goetz, J., Kiesler, S., & Powers, A. (2003). Matching robot appearance and behavior to tasks to improve human–robot cooperation. *Proceedings of the 12th IEEE International Workshop on Robot and Human Interactive Communication* (pp. 55–60). IEEE.

Gray, H. M., Gray, K., & Wegner, D. M. (2007). Dimensions of mind perception. *Science*, *315*(5812), 619–619. https://doi.org/10.1126/science.1134475

Haring, K. S., Silvera-Tawil, D., Takahashi, T., Watanabe, K., & Velonaki, M. (2016). How people perceive different robot types: A direct comparison of an android, humanoid, and non-biomimetic robot. *2016 8th International Conference on Knowledge and Smart Technology (KST)* (pp. 265–270). IEEE.

Hegel, F., Krach, S., Kircher, T., Wrede, B., & Sagerer, G. (2008). Understanding social robots: A user study on anthropomorphism. *The 17th IEEE International Symposium on Robot and Human Interactive Communication RO-MAN 2008* (pp. 574–579). IEEE.

Hoffner, C., & Buchanan, M. (2005). Young adults' wishful identification with television characters: The role of perceived similarity and character attributes. *Media Psychology*, *7*(4), 325–351. https://doi.org/10.1207/S1532785XMEP0704_2

Juarez, A., Bartneck, C., & Feijs, L. (2011). Using semantic technologies to describe robotic embodiments. *Proceedings of the 6th International Conference on Human–Robot Interaction* (pp. 425–432). ACM.

Judd, C. M., Westfall, J., & Kenny, D. A. (2012). Treating stimuli as a random factor in social psychology: A new and comprehensive solution to a pervasive but largely ignored problem. *Journal of Personality and Social Psychology*, *103*(1), 54–69. https://doi.org/10.1037/a0028347

Kalegina, A., Schroeder, G., Allchin, A., Berlin, K., & Cakmak, M. (2018). Characterizing the Design Space of Rendered Robot Faces. *Proceedings of the 2018 ACM/IEEE International Conference on Human–Robot Interaction* (pp. 96–104). ACM.

Kamide, H., Eyssel, F., & Arai, T. (2013). Psychological anthropomorphism of robots. *International Conference on Social Robotics* (pp. 199–208). Springer.

Kamide, H., Takubo, T., Ohara, K., Mae, Y., & Arai, T. (2014). Impressions of humanoids: The development of a measure for evaluating a humanoid. *International Journal of Social Robotics*, *6*(1), 33–44. https://doi.org/10.1007/s12369-013-0187-x

Kanda, T., Hirano, T., Eaton, D., & Ishiguro, H. (2004). Interactive robots as social partners and peer tutors for children: A field trial. *Human–Computer Interaction*, *19*(1), 61–84. https://doi.org/10.1207/s15327051hci1901&2_4

Katagiri, Y., Nass, C., & Takeuchi, Y. (2001). Cross-cultural studies of the computers are social actors paradigm: The case of reciprocity. In M. J. Smith, G. Salvendy, D. Harris, & R. Koubek (Eds.), *Usability evaluation and interface design: Cognitive engineering, intelligent agents, and virtual reality* (pp. 1558-1562). Lawrence Erlbaum Associates.

Kidd, C. D., Taggart, W., & Turkle, S. (2006). A sociable robot to encourage social interaction among the elderly. *Proceedings 2006 IEEE International Conference on Robotics and Automation, 2006. ICRA 2006* (pp. 3972–3976). https://doi.org/10.1109/ROBOT.2006.1642311

Koh, Y. J., & Sundar, S. S. (2010). Effects of specialization in computers, web sites, and web agents on e-commerce trust. *International Journal of Human–Computer Studies*, *68*(12), 899–912. https://doi.org/10.1016/j.ijhcs.2010.08.002

Kozima, H., Nakagawa, C., & Yano, H. (2003). Attention coupling as a prerequisite for social interaction. *The 12th IEEE International Workshop on Robot and Human Interactive Communication, 2003, ROMAN 2003* (pp. 109–114). https://doi.org/10.1109/ROMAN.2003.1251814

Lakatos, G., Gácsi, M., Konok, V., Brúder, I., Bereczky, B., Korondi, P., & Miklósi, Á. (2014). Emotion attribution to a non-humanoid robot in different social situations. *PLOS ONE*, *9*(12), Article e114207. https://doi.org/10.1371/journal.pone.0114207

Lee, E. J., Nass, C., & Brave, S. (2000). Can computer-generated speech have gender?: An experimental test of gender stereotype. *CHI'00 extended abstracts on human factors in computing systems* (pp. 289–290). ACM. https://doi.org/10.1145/633292.633461

Lee, K. M., Peng, W., Jin, S.-A., & Yan, C. (2006). Can robots manifest personality?: An empirical test of personality recognition, social responses, and social presence in human–robot interaction. *Journal of Communication*, *56*(4), 754–772. https://doi.org/10.1111/j.1460-2466.2006.00318.x

Lehmann, H., Saez-Pons, J., Syrdal, D. S., & Dautenhahn, K. (2015). In good company? Perception of movement synchrony of a non-anthropomorphic robot. *PLOS ONE*, *10*(5), Article e0127747. https://doi.org/10.1371/journal.pone.0127747

Leite, I., Martinho, C., & Paiva, A. (2013). Social robots for long-term interaction: A survey. *International Journal of Social Robotics*, *5*(2), 291–308. https://doi.org/10.1007/s12369-013-0178-y

Li, D., Rau, P. P., & Li, Y. (2010). A cross-cultural study: Effect of robot appearance and task. *International Journal of Social Robotics*, *2*(2), 175–186. https://doi.org/10.1007/s12369-010-0056-9

Li, J. J., Ju, W., & Reeves, B. (2017). Touching a mechanical body: Tactile contact with body parts of a humanoid robot is physiologically arousing. *Journal of Human–Robot Interaction*, *6*(3), 118–130. https://doi.org/10.5898/JHRI.6.3.Li

Lohse, M., Hegel, F., Swadzba, A., Rohlfing, K., Wachsmuth, S., & Wrede, B. (2007). What can I do for you? Appearance and application of robots. *Proceedings of AISB*, *7*, 121–126.

Mathur, M. B., & Reichling, D. B. (2016). Navigating a social world with robot partners: A quantitative cartography of the Uncanny Valley. *Cognition*, *146*, 22–32. https://doi.org/10.1016/j.cognition.2015.09.008

Miklósi, A., & Gácsi, M. (2012). On the utilization of social animals as a model for social robotics. *Frontiers in Psychology*, *3*, Article 75. https://doi.org/10.3389/fpsyg.2012.00075

Moon, J. W., & Kim, Y. G. (2001). Extending the TAM for a World-Wide-Web context. *Information & Management*, *38*(4), 217–230. https://doi.org/10.1016/S0378-7206(00)00061-6

Nass, C. (2004). Etiquette equality: Exhibitions and expectations of computer politeness. *Communications of the ACM*, *47*(4), 35–37. https://doi.org/10.1145/975817.975841

Nass, C., Fogg, B. J., & Moon, Y. (1996). Can computers be teammates? *International Journal of Human–Computer Studies*, *45*(6), 669–678. https://doi.org/10.1006/ijhc.1996.0073

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*(3), 171–181. https://doi.org/10.1037/1076-898X.7.3.171

Phillips, E., Ullman, D., de Graaf, M. M., & Malle, B. F. (2017). What does a robot look like?: A multi-site examination of user expectations about robot appearance. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* (Vol. *61*, pp. 1215–1219). https://doi.org/10.1177/1541931213601786

Phillips, E., Zhao, X., Ullman, D., & Malle, B. F. (2018). What is human-like?: Decomposing robots' human-like appearance using the Anthropomorphic roBOT (ABOT) database. *Proceedings of the 2018 ACM/IEEE International Conference on Human–Robot Interaction* (pp. 105–113). ACM.

Powers, A., & Kiesler, S. (2006). The advisor robot: Tracing people's mental model from a robot's physical attributes. *Proceedings of the 1st ACM SIGCHI/SIGART Conference on Human–Robot Interaction* (pp. 218–225).

Reeves, B., Hancock, J., & Liu, X. S. (2019). *Social robots are like real people: First impressions, attributes, and stereotyping of social robots*. https://goo.gl/Gqpzkx; https://goo.gl/eejbV7

Reeves, B., & Nass, C. I. (1996). *The media equation: How people treat computers, television, and new media like real people and places*. Cambridge University Press.

Russell, A. M. T., & Fiske, S. T. (2008). It's all relative: Competition and status drive interpersonal perception. *European Journal of Social Psychology*, *38*(7), 1193–1201. https://doi.org/10.1002/ejsp.539

Sabanovic, S., Michalowski, M. P., & Simmons, R. (2006). Robots in the wild: Observing human–robot social interaction outside the lab. *9th IEEE International Workshop on Advanced Motion Control, 2006* (pp. 596–601). https://doi.org/10.1109/AMC.2006.1631758

Salganik, M. (2019). *Bit by bit: Social research in the digital age*. Princeton University Press.

Satake, S., Kanda, T., Glas, D. F., Imai, M., Ishiguro, H., & Hagita, N. (2009). How to approach humans?: Strategies for social robots to initiate interaction. *Proceedings of the 4th ACM/IEEE International Conference on Human Robot Interaction* (pp. 109–116). ACM.

Todorov, A., Said, C. P., Engell, A. D., & Oosterhof, N. N. (2008). Understanding evaluation of faces on social dimensions. *Trends in Cognitive Sciences*, *12*(12), 455–460. https://doi.org/10.1016/j.tics.2008.10.001

Ueyama, Y. (2015). A bayesian model of the uncanny valley effect for explaining the effects of therapeutic robots in autism spectrum disorder. *PLOS ONE*, *10*(9), Article e0138642. https://doi.org/10.1371/journal.pone.0138642

von der Pütten, A. M., & Krämer, N. C. (2012). A survey on robot appearances. *Proceedings of the Seventh Annual ACM/IEEE International Conference on Human–Robot Interaction* (pp. 267–268). ACM.

Walters, M. L., Koay, K. L., Syrdal, D. S., Dautenhahn, K., & Te Boekhorst, R. (2009). Preferences and perceptions of robot appearance and embodiment in human–robot interaction trials. Adaptive and emergent behavior and complex systems. *Proceedings of the 23rd Convertion of the society for the study of artificial intelligence and simulation of behavior, AISB 2009* (pp. 136-143).

Willis, J., & Todorov, A. (2006). First impressions: Making up your mind after a 100-ms exposure to a face. *Psychological Science*, *17*(7), 592–598. https://doi.org/10.1111/j.1467-9280.2006.01750.x

Yanco, H. A., & Drury, J. (2004). Classifying human–robot interaction: An updated taxonomy. *2004 IEEE International Conference on Systems* (Vol. *3*, pp. 2841–2846). IEEE.