Robot or Human?

Consumer Perceptions of Human-Like Robots

NOAH CASTELO

BERND SCHMITT

MIKLOS SARVARY

NADIA THALMANN

Noah Castelo (nc2609@columbia.edu) is a PhD candidate at Columbia University, New York, NY, 10025. Bernd Schmitt (bhs1@gsb.columbia.edu) is the Robert D. Calkins Professor of International Business at Columbia University, New York, NY, 10025. Miklos Sarvary (miklos.sarvary@columbia.edu) is the Carson Family Professor of Business at Columbia University, New York, NY, 10025. Nadia Thalmann (nadiathalmann@ntu.edu.sg) is Professor of Computer Science and Engineering at Nanyang Technological University, Singapore. Correspondence: Noah Castelo. The authors are grateful for the assistance of Geoffrey Heal in preparing the stimuli used in this research.

This paper is being actively revised for second round review at the Journal of Consumer Research. Please do not circulate without permission.

This version: August 30, 2018.

CONTRIBUTION STATEMENT

Social robots (intended to interact directly with consumers) are the fastest growing segment of the robotics market, and yet consumer research on such robots is almost non-existent. This paper studies the central question of how consumers perceive the kinds of social robots that are being increasingly used in retail and service settings and being marketed directly to consumers for use in the home, and how these perceptions affect consumers' evaluations of such robots and the firms employing them. We show that the degree of a robot's physical human-likeness is a key driver of consumer perceptions of social robots. This finding contributes directly to existing research on the uncanny valley phenomenon. Most importantly, we move significantly beyond this existing research by focusing on consumer perceptions of highly and even perfectly human-like robots, which will be available to consumers and businesses in the near future and which no prior research to our knowledge has studied. We propose a novel theory based on mind perception in robots to explain why perfectly human-like robots are evaluated more positively than highly human-like robots, but more negatively than actual humans. We believe that our research will stimulate a great deal of new theoretical and empirical work on the important and understudied topic of robots in consumer settings.

ABSTRACT

Robots intended to interact directly with humans (social robots) are the fastest growing segment of the robotics market, and are being employed in retail stores, restaurants, hospitals, hotels, and homes. The effect that these robots have on the consumer experience depends largely on how human the robots look. Using pictures and videos of real consumer robots interacting with humans as stimuli, we show that the effect of robots' physical human-likeness on consumers' reaction to social robots reverses from negative to positive as human-likeness approaches perfection. Moreover, we show that perfectly human-like robots are evaluated more positively than highly human-like robots, but more negatively than actual humans. We develop and test a theory based on mind perception in robots to explain these novel findings. Overall, our results suggest that consumers are uncomfortable with robots whose appearance blurs the line between human and machine and point to specific practical approaches for reducing this discomfort.

Keywords: *technology, robotics, consumer experience, cross-cultural research.*

Robots are becoming increasingly prevalent and important in many domains of business and consumer behavior. Today, there are already more than 1.5 million robots used worldwide in manufacturing alone. However, recently the market for social robots (which are intended to interact directly with consumers at home and in retail and service contexts) has been growing seven times faster than the market for manufacturing robots (Business Insider 2015), reaching $5.4 billion in sales by the end of 2017 and expected to triple to $14.9 billion within the next five years (Business Wire 2017). Consumers can already purchase social robots that perform chores, engage and monitor young and elderly people, engage in conversations, and act as companions and assistants (Gibbs 2016). Outside the home, social robots are being used in a wide variety of contexts including retail stores, restaurants, hotels, and hospitals (Dass 2017; Nguyen 2016; Simon 2015).

Social robots vary significantly in terms of how human-like they look. Some social robots, such as Jibo (sold for use in the home and recently featured on the cover of *Time* magazine as one of the best inventions of 2017) allude to human features by using a large round screen for a face but lack any recognizably human body. Other robots such as Pepper (a robot used in retail stores and restaurants) are more humanoid in their appearance, having a more human-like face as well as extremities resembling human legs, arms, and hands. Finally, some robots such as Sophia (who appears as a speaker at conferences and as a guest on talk shows) and Erica (recently employed as a hotel concierge and news anchor in Japan) are designed to be exact replicas of humans and are increasingly difficult to distinguish from real humans. In fact, rapid ongoing progress in robotics and artificial intelligence (AI) research is pushing robots ever closer to human-level appearance and intelligence (Goertzel 2014; Goertzel, Hanson, and Yu 2014; Ishiguro and Nishio 2007). Human-level physical appearance is quickly approaching, and experts estimate that AI will have a 50% chance of achieving human-level intelligence within the next 20-25 years (Müller and Bostrom 2016).

In a world in which social robots are entering the consumer marketplace and rapidly approaching perfect human-likeness, it is imperative for consumer researchers to study how consumers might experience and evaluate social robots, as well as the firms that employ and sell them, in various home and business settings. Thus, the following consumer questions arise: How do consumers react to robots of differing human-likeness? What factors determine those reactions? Finally, and most intriguingly, might consumers perceive a human-like mind in robots with human-like appearances, and how might mind perception impact evaluations of such robots?

In this paper, we address these questions in six studies. We provide evidence that increasing human-likeness has initially negative effects on consumer evaluations of social robots. However, beyond a certain point of human-likeness, as we approach perfectly human-like robots, further increases in human-likeness have positive effects; yet even perfectly human-like robots elicit more negative reactions than real human beings—a novel finding that we explain with a theory based on the perception of a mind in human-like robots. Furthermore, we show that the perception of a mind in a robot is malleable and varies across cultures, and that increasing the perception of mind improves consumers' evaluations of highly human-like robots. Finally, building on the literature in mind perception, we show that a robot must be seen as having both dimensions of mind (agency and experience) in order for these positive effects of mind perception to occur.

These results contribute to the literatures on human-robot interaction and mind perception by explaining consumers' perceptions of very human-like and even perfectly human-like robots and documenting for the first time a theory-driven approach for improving evaluations of very human-like robots. We also contribute to the marketing literature by demonstrating how firms that employ such robots can improve their consumers' experiences and ultimately the firm itself is perceived and evaluated. Our findings therefore also have practical implications for the production and marketing of social robots.

THE UNCANNY VALLEY

Prior research on the relationship between robots' human-likeness and human evaluation of robots has been guided by the "uncanny valley hypothesis." Nearly 50 years ago, Masahiro Mori, a Japanese roboticist, wrote an influential paper speculating that making robots look more human-like is beneficial only up to a point, after which they become *too* human-like and elicit strongly negative reactions (Mori 1970). The "valley" thus refers to the worsening of reactions as robots move from moderately human-like to very (but not perfectly) human-like, and the subsequent improving of responses as human-likeness approaches perfection (i.e., actual humans). Mori therefore advised that social robots should be designed to have a moderate degree of human likeness in order to avoid the negative reactions elicited by very human-like robots.

In the years since Mori's paper, research has produced inconsistent findings regarding whether the uncanny valley exists in the form originally proposed. For example, while some studies have found support for a non-linear relationship between a robot's human-likeness and affective reactions to the robot (Ferrari, Paladino, and Jetten 2016; Mathur and Reichling 2016), several others have found no such pattern (Bartneck et al. 2009; Rosenthal-Von Der Pütten and Krämer 2014; Zlotowski, Proudfoot, and Bartneck 2013). Indeed, two recent reviews of research on the uncanny valley concluded that “although the notion of the uncanny valley is plausible and is supported by plentiful anecdotal evidence, rigorous controlled studies have yielded mixed support for its existence” (Wang, Lilienfeld, and Rochat 2015) and “it is surprising that empirical evidence for the uncanny valley hypothesis is still ambiguous if not non-existent” (Kätsyri et al. 2015).

These inconsistent findings can likely be explained by the fact that Mori’s original hypothesis was presented as a broad, somewhat vague idea without precisely defined constructs and without any data. This inevitably resulted in subsequent research operationalizing the two key variables – the robots’ human-likeness and human reactions to the robots – inconsistently. For example, the original hypothesis operationalized “reactions” in terms of feelings of eeriness and creepiness (Mori 1970), while more recent research has measured feelings of unease (Gray and Wegner 2012), likeability of the robot (Mathur and Reichling 2016), and repulsion (Ferrari et al. 2016). Our main interest in this research is consumers’ comfort interacting with robots, which we argue (and demonstrate) is a key variable in shaping a range of consequential marketing outcomes. Specifically, consumers’ comfort interacting with robots impacts their likelihood of shopping in a store or visiting a hospital where robots are employed and their evaluations of companies that employ robots. In addition to measuring comfort directly, we also use measures of “uncanniness” employed in prior research on the uncanny valley, which is closely related to comfort and to the original formulation of the uncanny valley hypothesis (Gray and Wegner 2012).

Importantly, no existing research has explored reactions to perfectly human-like robots, and the original formulation of the uncanny valley hypothesis made no explicit predictions about such robots. From a consumer point of view, exploring reactions to perfectly human-like robots is important because the pace of technological progress suggests that consumers will soon be interacting with robots whose physical appearance (and potentially even whose intelligence) may be indistinguishable from humans in many commercial and social settings. If a robot is indistinguishable from a human in appearance and/or behavior, then there seem to be two possible consumer reactions, depending on whether consumers are aware or not aware of whether the entity they are interacting with is a robot or a human. If they are unaware that the robot is a robot, then reactions to the robot should be as positive as reactions to a real human. On the other hand, if people know that the perfectly human robot is in fact a robot, then they may still exhibit negative reactions because perfectly human robots may evoke certain concerns such as threats to human jobs, to human distinctiveness and even to human safety (Ferrari et al. 2016; Rosenthal-Von Der Pütten and Krämer 2014), and, most importantly, because even perfectly human-like robots may not be perceived to have a mind in the same way that humans do.

In this paper, we provide evidence consistent with the original formulation of the uncanny valley hypothesis that consumer reactions to robots change as robots move "into" the uncanny valley (i.e., as the human-likeness of robots increases from moderate to high) and "out of" the uncanny valley (i.e., as their human-likeness further increases from high to perfect). There are several existing theories that can be used to explain the into the valley effect, which we review below. We will show, however, that a new theory based on mind perception provides a better explanation of the out of the valley effect.

INTO THE VALLEY: THREATS AND AESTHETICS

Some existing research has shown that reactions to robots worsen as the robot moves from moderate to high human likeness – that is, as they move into the uncanny valley (Ferrari et al. 2016; Mathur and Reichling 2016). There are several theories that can potentially explain this finding.

Threat-based theories of uncanniness

Human-like robots pose *symbolic threats*, which are threats to beliefs, values, or worldviews (Riek, Mania, and Gaertner 2006). For example, the *threat to human distinctiveness hypothesis* explains negative reactions by proposing that highly human-like robots call into question the clear distinction between humans and machines, and that this threatens the fundamental desire to feel that one is a member of a distinct group (Ferrari et al. 2016). Thus, increasing human-likeness may increase symbolic threats to humans, in turn creating negative reactions towards more human-like robots.

Furthermore, human-like robots pose *realistic threats* to humans. *Realistic group conflict theory* proposes that when two groups are competing for resources, the potential success of one group threatens members of the other group, resulting in animosity between the groups (Jackson 1993). Although this theory has not been applied specifically to the context of robots, it seems that humans will be competing against robots for employment because robots will perform some of the same roles that humans currently perform. As a result, robots will pose a realistic threat: a recent report found that up to 38% of U.S. jobs will be replaced by machines in the next 15 years, or about 800 million jobs (PwC 2017). Increasing human-likeness may increase the degree to which a robot is seen as posing a perceived threat to human jobs as more human-like robots may seem more capable and less distinct from humans, thereby leading to more negative reactions toward the robots as human-likeness increases.

A second form of realistic threat concerns robots becoming physically dangerous to humans, potentially even threatening the survival of the human species. This theme is often featured in science fiction, such as in the films *2001: A Space Odyssey, Ex Machina,* and the television shows *Westworld* and *Humans*, and has even been promoted by prominent academics and businesspeople. For example, Stephen Hawking, the British theoretical physicist, warned that AI could "spell the end of mankind" (Rory Cellan-Jones 2014), and a report from Oxford University's Future of Humanity Institute ranked AI as the greatest risk to human extinction (Sandberg and Bostrom 2008). Increasing human-likeness may also increase the perceived threat from robots to human safety, thereby also worsening reactions to robots.

Aesthetic-based theories of uncanniness

In contrast to the threat-based explanations of uncanniness, other theories focus on how the robots' physical or aesthetic appearance contributes to reactions to robots, independent of perceived threats. The *violation of expectation hypothesis* explains that human-like robots elicit negative reactions because they create expectations of a human being, but something about their appearance fails to meet those expectations (Mitchell et al. 2011).

Finally, the *evolutionary aesthetics hypothesis* explains uncanny reactions to human-like robots by suggesting that evolution has shaped human preference for physical appearances that signal health and fitness, and that humanoid robots lack such an appearance because of their aesthetic imperfections (Hanson 2005). Thus, any physical imperfection that signals potential ill health or increases the salience of disease or death (not only among robots but among humans as well) can produce feelings of unease and uncanniness.

Each of these potential explanations can be used to explain the existing finding (the into the valley effect) that highly but imperfectly human-like robots elicit feelings of discomfort and creepiness among consumers. Applied to the specific case of perfectly human-like robots, they have conflicting predictions. Threat-based theories would predict that perfectly human-like robots should be seen as more threatening, in turn causing reactions to worsen even beyond reactions to existing humanoid robots, whose appearance is highly (but not perfectly) human-like. Aesthetic-based theories, on the other hand, would predict that reactions to perfectly human-like robots would improve compared to existing humanoid robots and that they would be the same as for humans. An improvement in comfort from high to perfect human-likeness would therefore suggest that aesthetics have a stronger effect than perceived threats on consumers' comfort with robots. In contrast, we expect that aesthetics will have a weaker effect than perceived threats, but that comfort will still improve as human-likeness approaches perfection. The reason we posit for this improvement is that robots with very high human-likeness will be perceived to have a human-like mind.

OUT OF THE VALLEY: MIND PERCEPTION

As robots' human-likeness increases beyond anything that has been studied thus far (but that is already being created by roboticists), we propose that a new question becomes relevant, namely: could this robot have a human-like mind? Specifically, we propose a positive relationship between a robot's physical human-likeness and the degree to which it is seen as having a human-like mind. In this research, we define and measure mind both as a single intuitive concept as well as in terms of the two dimensions of mind that prior research has identified: agency (the ability to plan and act autonomously) and experience (the capacity to feel emotions and sensations; Gray, Gray, and Wegner 2007; Waytz et al. 2010). We expect that overall perceptions of "mind" as well as of the two constituent dimensions of mind will increase with physical human-likeness.

Furthermore, we expect that perceiving a mind in a robot will lead to more positive evaluations of the robot, because consumers will feel empathy for robots with a mind – i.e., the ability to understand and share the robots' thoughts and feelings – which will outweigh the potential threats posed by such robots. However, we expect that mind perception will have these positive effects only among robots with high human-likeness, since only such robots will have "enough of a mind" to outweigh the otherwise threatening aspects of human-like robots. In other words, robots with very low human-likeness may be seen as having a very low degree of mind, but not enough to create positive evaluations of the robot, similar to how animals are granted some degree of mind (i.e., some capacity to experience emotion and act autonomously), but are still killed and eaten. The precise level of human-likeness at which mind perception becomes salient enough to trigger this process is unclear, since category membership is fuzzy and continuous (Medin and Smith 1984). We will therefore identify the threshold empirically in our studies.

However, despite any improvement in comfort that might occur as robots move from highly to perfectly human-like, we further predict that even perfectly human-like robots will still likely elicit more negative reactions than a real human would elicit. That is, consumers will be less likely to grant a full mind to robots even if they are given every indication that the robot has *all* the same capacities and appearances as a real human – merely because they are a robot. The robot could be perfectly human-like in every sense – not only in terms of physical appearance, but also in its movement, speech, and conversational abilities – and still not be perceived as positively as real humans because humans will refuse to grant a complete mind to a non-human species. Formally, our hypotheses are as follows:

H1: Consumer evaluations of robots (i.e., comfort interacting with robots, perceived uncanniness of robots) will worsen as human likeness increases from low to high ("into the valley").

H2: Consumer evaluations of robots will improve as human likeness increases from high to perfect but remain worse than evaluations of humans ("out of the valley").

H3: The effects predicted in H2 will occur because of *mind perception*: more human-like robots will be seen as having more of a mind, which has positive effects on evaluations of highly human-like robots, but perfectly human robots will still be seen as having less of a mind than humans.

H4: Mind perception will have positive effects on evaluations because it triggers empathy.

We also expect that these effects will have practical marketing implications. As more businesses begin to consider the use of robots in their stores, a key question is what kind of robots should be used and how the robots will impact consumers' experience in their stores. We suggest that consumers will be less comfortable shopping in a store, staying in a hotel, or visiting a hospital where robots are employed if those robots make consumers feel uncomfortable. Similarly, we expect that this discomfort will generalize to include the company employing the robot, leading consumers to negatively evaluate the company.

H5: Consumer evaluations of robots will in turn impact marketing outcomes for firms selling and employing robots (i.e., shopping in stores where robots are employed, attitudes towards companies employing robots).

OVERVIEW OF THE STUDIES

Our initial studies will explore how consumer reactions to social robots varies as human-likeness increases from low to high to perfect, using both videos (Study 1) and pictures of social robots (Study 2). Study 1 shows that the effect of human-likeness on consumer reactions does indeed reverse (from negative to positive) as human-likeness increases. Study 2 replicates this pattern with a broader set of stimuli and shows that mind perception plays a significant role in explaining the out of the valley phenomenon. Study 3 shows that the effects of mind perception on evaluations of highly human-like robots are stronger than the effects of perceived threats and aesthetics. Study 4 provides causal evidence for the role of mind perception, showing that portraying the mind as a purely physical process that can be replicated in machines improves consumers' evaluations of robots and downstream marketing outcomes associated with their use. Study 5 provides a cross-cultural context in which to test the effect of mind perception on evaluations of robots: compared to Americans, Japanese consumers are more likely to perceive a mind in inanimate objects, and this tendency in turn explains more positive evaluations of robots in Japan. Finally, Study 6 shows that both dimensions of mind – agency and experience – are required in order for these positive effects of mind perception on evaluations of robots to occur.

STUDY 1

Our first study uses videos of robots that vary in human-likeness in order to explore the effects of human-likeness on consumer evaluations of robots. This study tests H1, the hypothesis that increases in human-likeness will have initially negative and then positive effects on consumer evaluations.

Method

200 American participants recruited from MTurk (55% female, mean age = 37) watched one of three videos that we produced depicting a robot (or a human) interacting with a human using a standardized script. The videos featured either Nao (a moderately human-like robot), Nadine (a highly human-like robot), or the human being whom Nadine was designed to resemble (links to the videos are provided in the Appendix). Participants who saw the human video were either told that the human was in fact a highly advanced robot (human-as-robot condition), or that she was indeed a human (human-as-human condition). Note that the video was exactly the same in these two conditions, and the human actor was blind to our hypotheses and instructed to act naturally, following the same script as the other two videos that featured the actual robots. In order to eliminate as many differences between the conditions as possible, each video was cropped to show only the robot's (or the human's) head and shoulders. After watching the video, participants completed a 3-item measure of their evaluation of the robot or human, taken from prior research on the uncanny valley phenomenon (how much would you feel *uneasy/unnerved/creeped out* during an interaction with this robot/person, on a 0–10 scale anchored at "not at all" and "completely," $\alpha = .95$; Gray and Wegner 2012). This was the only dependent variable measured in this study. We also asked participants whether anything about

the study seemed suspicious and excluded 16 participants (30%) in the human-as-robot condition who reported suspicion that the human was in fact a robot, although the pattern of results is unchanged if we include these participants.

Results and Discussion

An ANOVA revealed that evaluations varied significantly across the videos, $F(1,180) = 14.55$, $p < .001$ (see Figure 2). Participants' evaluations initially worsened as human-likeness increased from moderate ($M_{moderate} = 2.93$) to high ($M_{high} = 5.58$, $t(94) = 5.11$, $p < .001$), and then improved as human-likeness increased from high to perfect ($M_{perfect} = 3.57$, $t(70) = 3.22$, $p = .002$ compared to M_{high}; recall that lower scores on this measure reflect more positive evaluations). However, the perfect robot was still rated more negatively than the human portrayed as a human ($M_{human} = 2.32$, $t(86) = 2.13$, $p = .037$). This study therefore provides initial evidence for the distinct "into the valley" and "out of the valley" effects as human-likeness increases.

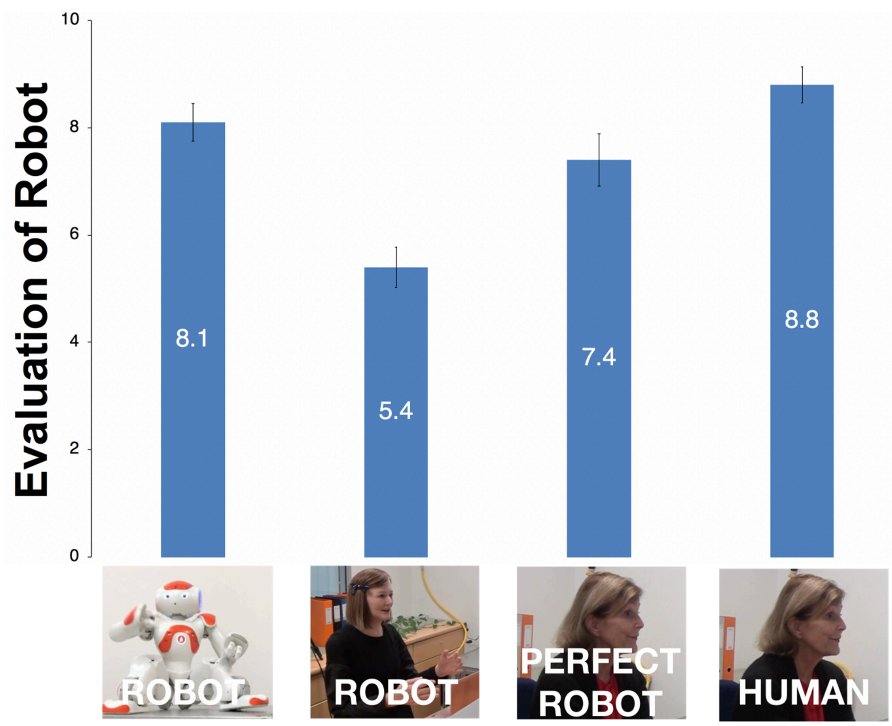


Figure 2. Results of Study 1. Error bars represent standard errors.

The into the valley effect in this study is consistent with the threat- and aesthetic-based theories described earlier: more human-like robots may be more threatening to human jobs, safety, and distinctiveness, and may also look physically unpleasant or low-quality. Interpreting the out of the valley effect in terms of these explanations, however, is problematic: the improving evaluations from the highly to perfectly human-like robot suggests that aesthetics dominate threats, whereas the gap between the perfectly human-like robot and the actual human suggests that threats dominate aesthetics. The following studies will show instead that mind perception provides a more consistent explanation for this pattern of results.

An initial marketing implication of these results is that firms may be better off employing robots with low or moderate human-likeness, which in this case was evaluated as positively as a human ($t(100) = 1.29$, $p = .209$) and as positively as a perfectly human-like robot ($t(91) = 1.07$, $p = .289$). Thus, despite the higher costs associated with creating perfectly human-like robots, these robots are perceived no more positively than moderately human-like robots, and more negatively than actual humans.

STUDY 2

The second study is intended to replicate the pattern of results seen in Study 1 using a wider range of stimuli. In addition, Study 2 will provide initial evidence for mind perception as an explanation of these results. This study employed a larger dataset of real social robots that vary significantly in terms of physical human-likeness as stimuli, which decreases the likelihood that our results thus far were driven by idiosyncratic individual stimuli.

Method

We recruited 800 American participants (51% female, mean age = 33) from Prolific Academic, a crowdsourcing website where participants are more naïve to common experimental paradigms and more honest than participants on Mechanical Turk (Peer et al. 2017). Participants were shown images of robots compiled by the Anthropomorphic Robot Database (Phillips et al. 2018). We chose 25 of these robots varying in human-likeness from very low to very high. 200 of our participants rated the robots on how human-like they looked overall. This allowed us to divide the 25 robots into five quintiles based on their overall human-likeness. The remaining 600 participants were shown five of the robots (one from each of the five quintiles) or one of four human beings. Participants were either told that these humans were humans or were in fact advanced humanoid robots. This allows us to compare evaluations of perfectly human-like robots to evaluations of actual humans. Please see appendix for exact stimuli used in all studies. As our primary dependent variable, we asked participants how creepy each robot seemed and how comfortable they would be interacting with each robot. We also asked how much each robot (or human) seemed like it could "have a mind like humans do" (as a measure of overall mind perception), "experience emotion like humans do," and "have autonomy like humans do" (as measures of experience and agency). We chose to measure emotion specifically as our measure of "experience," because the sensation aspect of experience as it is typically defined is less relevant for our purposes given that all robots have sensation (i.e., vision). Each item used 0–10 scales anchored at "not at all" and "completely." We excluded 99 participants (16.5%) who reported suspicion that the humans portrayed as robots were not in fact robots, although the pattern of results in unchanged if we include these participants.

Results and Discussion

In order to test for a U-shaped relationship between human-likeness and consumer reactions, we conducted a two-lines test, which is a more accurate and valid method of testing for such relationships than testing for a quadratic term in a regression (Simonsohn 2018). This test estimates separate regression lines for low and high values of the x-variable (using an algorithm to determine the boundary point between low and high that maximizes the overall statistical power of the test) and is significant if the two lines are individually significant and opposite in sign. The results of this test are depicted in Figure 3 and show that increasing human-likeness leads to worse reactions until the point 8.14 (slightly lower than the human-likeness of the most human-like robots that currently exist; $\beta = -.52$, $p < .001$). After this point, further increases in human-likeness lead to improved reactions, $\beta = 1.89$, $p = < .001$). Note that this analysis did not include the humans portrayed as humans.

It is worth noting that we do not observe the initial improvement in evaluations as human-likeness moves from low to moderate that Mori predicted when he proposed the uncanny valley hypothesis. The reason for this may be that Mori included industrial robots in his discussion, which have extremely low human-likeness, whereas our stimuli included only social robots that have at least some subtle allusions to human-likeness, such as a large round screen that alludes to a face. The initial improvement that Mori predicted may therefore only occur when non-human-like industrial robots are included as stimuli.

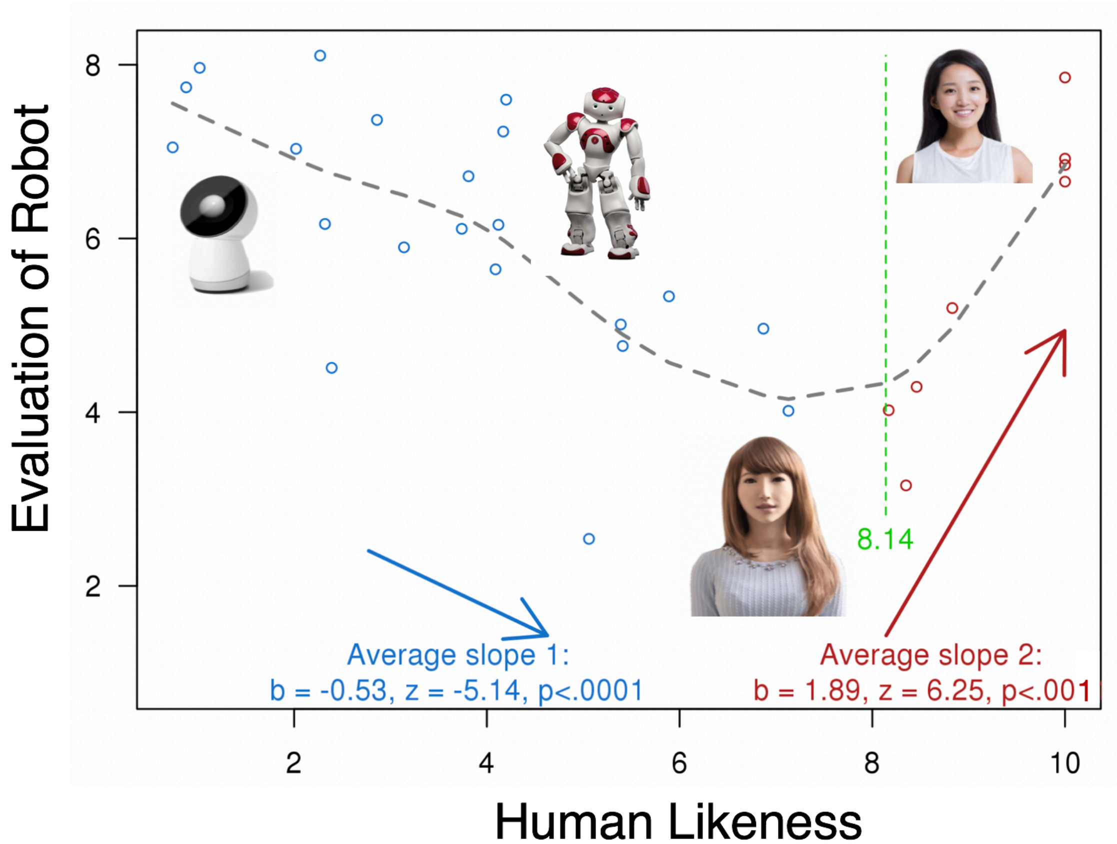


Figure 3. Results of Study 2. Human-likeness first elicits more negative reactions, then more positive reactions. Individual points on the graph represent the individual robots used as stimuli.

The perfectly human-like robots in this study, as in Study 1, were evaluated more negatively than actual humans ($M_{Human} = 8.70$ vs. $M_{PerfectRobot} = 6.87$, $t(503) = 9.65$, $p < .001$), but more positively than the robots in the fifth (highest) quartile of human likeness ($M_{Q5} = 4.16$, $t(553) = 11.64$, $p < .001$), as well as the robots in the fourth quartile ($M_{Q4} = 4.84$, $t(416) = 7.39$, $p < .001$), third quartile ($M_{Q3} = 6.17$, $t(416) = 2.72$, $p = .007$), and second quartile ($M_{Q2} = 6.05$, $t(416) = 3.07$, $p = .002$), but more negatively than the robots in the first (lowest) quartile of human likeness ($M_{Q5} = 7.48$, $t(416) = -2.48$, $p = .013$).

We also found evidence that mind perception plays a role in explaining this pattern of results. Human likeness had a linear positive relationship with perceived mind in the robots, $\beta = .22$, $p < .001$, and perceived mind in turn had a non-linear relationship with evaluations of the robots. Perceived mind and human-likeness had a significant interaction on participants' evaluations of robots, $\beta = .22$, $p = .012$. Among robots with human-likeness ratings less than 8.14 (the dividing line between "into" and "out of" the uncanny valley determined by the two-lines test), perceived mind had a non-significant negative effect on evaluations, $\beta = -.32$, $p = .489$. In contrast, among robots with human-likeness ratings above 8.14, perceived mind had a positive effect on evaluations, $\beta = 1.70$, $p = .059$.

Perfect robots were granted less of a mind than actual humans ($M_{Human} = 9.26$ vs. $M_{PerfectRobot} = 4.17$, $t(503) = 23.52$, $p < .001$), and the difference in overall evaluations of humans and perfect robots was predicted by the difference in mind attributed to humans and prefect robots, $\beta = .28$, $p < .001$.

We also found this same pattern of results using the measures of agency and experience instead of the overall measure of mind perception. These two measures were highly correlated, $r(598) = .86$, $p < .001$, and were both highly correlated with overall mind perception (agency $r(598) = .91$, $p < .001$; experience $r(598) = .92$, $p < .001$). Both measures individually and together (averaged) had the same interaction effect on evaluation of robots, so we report them together. The two-dimension measure of mind perception also interacted with human-likeness, $\beta = .28$, $p = .008$. Among robots with human-likeness ratings less than 8.14 (the dividing line between "into" and "out of" the uncanny valley determined by the two-lines test), perceived mind had a non-significant negative effect on evaluations, $\beta = -.79$, $p = .125$. In contrast, among robots with human-likeness ratings above 8.14, perceived mind had a positive effect on evaluations, $\beta = 1.89$, $p = .076$.

Existing research on the uncanny valley phenomenon has demonstrated the into the valley effect wherein evaluations become more negative as human-likeness increases from low to high and has explained this effect in terms of perceived threats and aesthetic factors (Ferrari et al. 2016; Mathur and Reichling 2016). The results of our first two studies replicate this into the valley effect and make a novel contribution by showing for the first time what happens when human-likeness approaches perfection. Importantly, mind perception plays a role in explaining this out of the valley effect.

This pattern also reinforces the marketing implication derived from Study 1, that firms who choose to employ social robots may be better off employing robots with low-to-moderate human-likeness. However, the robotics industry is clearly pushing towards robots with ever greater human-likeness, suggesting that some companies might consider employing and selling such robots. Our remaining studies therefore focus on testing our mind perception account of the out of the valley effect in greater detail, in order to shed light on how firms that produce, employ, and sell highly human-like robots can help bring those robots out of the uncanny valley.

STUDY 3

The purpose of this study is to show that mind perception is a better explanation of the out of the valley effect than perceived threats and aesthetic factors. Relying only on threat- and aesthetic-based theories to explain our pattern of results would suggest that perfectly human-like robots are evaluated more positively than highly human-like robots because they are more aesthetically appealing, whereas perfectly human-like robots are evaluated more negatively than actual humans because they are more threatening. Instead, our mind perception account provides a simpler explanation for the entire out of the valley effect, which this study will confirm empirically.

Method

155 American MTurk participants (37% female, mean age = 35) were assigned to one of three conditions. In the human condition, participants saw a picture of a human and were told that a chain of pharmacies was considering hiring this person. In the perfect robot condition, participants saw the same picture of a human but were told that it was an advanced humanoid robot and that the pharmacy was considering a plan to start using this kind of robot in their stores. In the robot condition, participants saw a picture of an actual social robot (Erica) and were told that the pharmacy was considering a plan to start using this kind of robot in their stores. The robots were described as being able to assist customers, answer questions, and work as cashiers We excluded 8 participants in the human-as-robot condition who did not believe that it was a robot.

Participants then reported how comfortable they would be shopping in a store that employed this kind of robot (or person). Participants also reported how much the robot or person they saw seemed to a mind of its own in order to test for the effects of mind perception. In order to test for these effects over and above threat- and aesthetic-based explanations, we also measured the perceived threat that the robot or person posed to jobs, safety, and distinctiveness, as well how unhealthy they looked (because perceived ill health is a key factor in aesthetic explanations of the uncanny valley). All measures used 0–10 scales.

Results

Participants expected to feel most comfortable shopping in a store staffed by the human (M = 8.44), less comfortable in a store staffed by a perfectly human-like robot (M = 5.56, $t(93) = 4.95$, $p < .001$), and least comfortable in a store staffed by a highly but imperfectly human-like robot (M = 4.35, $t(100) = 1.98$, $p = .050$ compared to the perfectly human-like robot).

The human also seemed to have the most mind (M = 8.86), while the perfectly human-like robot had less of a mind (M = 2.68, $t(93) = 11.12$, $p < .001$) and the highly but imperfectly human-like robot had the least (M = 1.69, $t(100) = 1.78$, $p = .078$ compared to the perfectly human-like robot). The perfectly human-like robot seemed to pose a greater threat to human jobs (M = 7.63) than the highly but imperfectly human-like robot (M = 5.67, $t(100) = 3.43$, $p < .001$), a greater threat to human distinctiveness (M = 5.34 vs. M = 3.38, $t(100) = 3.13$, $p = .002$), and looked less like an unhealthy person (M = 1.42 vs. M = 2.73, $t(100) = -2.46$, $p = .015$). However, perceived mind remained a highly significant and positive predictor of comfort ($\beta = .41$, $p < .001$) when controlling for perceived threats to jobs ($\beta = -.04$, $p = .601$), safety ($\beta = -.35$, $p < .001$), and human distinctiveness ($\beta = -.08$, $p = .348$), as well as how much the robot or human looked unhealthy (i.e., aesthetic appeal; $\beta = -.09$, $p = .294$). Thus, perceived mind is a stronger predictor of comfort shopping in a store staffed by a humanoid robot than any of these other potential predictors, and perceived mind mediated the effect of condition on comfort ($\beta = -2.46$, 95% CI [-3.8, -1.0].

These results therefore confirm that the existing theories used to explain the uncanny valley phenomena are incapable of explaining the novel out of the valley effect observed the context of highly human-like robots. Instead, mind perception provides a superior theoretical account of our findings.

STUDY 4

This study will provide causal evidence for the role of mind perception in shaping consumer evaluations of highly human-like robots by directly manipulating the degree to which robots are seen as being able to have a mind. In addition, this study measures evaluations of robots in specific commercial settings and evaluations of companies employing robots in order to strengthen the marketing implications of our research.

Method

100 American MTurk participants (46% female, mean age = 36) were assigned to watch one of two videos, featuring a professor explaining that robots either could or could not have a mind in the same way that humans do (links to the videos are provided in the Appendix). The arguments were taken from the philosophical literature on mind and consciousness, and mirrored the “dualism vs. physicalism” debate that portrays human mind as having non-material aspects that cannot be replicated in a machine (“dualism”) or as something explainable in terms of purely physical brain processes that can be replicated in a machine (“physicalism”) (Dennett 1994; Searle 1995).

After watching one of the videos, participants were asked to spend two minutes writing about why the arguments in the video were likely to be true, which is a technique known as the “saying-is-believing technique” commonly used in the psychology literature to increase engagement with stimuli and facilitate attitude change (Higgins and Rholes 1978; Aronson, Fried, and Good 2002; Yeager et al. 2016). We excluded 4 participants who failed to write a single comprehensible sentence. As a manipulation check, we asked participants whether they believed that robots could have a mind, on a 0 (not at all) to 10 (completely) scale.

All participants were then shown a picture of an advanced humanoid robot. The picture was accompanied by an explanation that humanoid robots are fast approaching perfect human-likeness and are being used as employees in stores, restaurants, and hotels, and that one such robot (Sophia) had even been granted citizenship in Saudi Arabia. As our primary dependent variables, we then asked participants how comfortable they would be shopping in a store and dining in a restaurant where this kind of robot was employed, and how interested they would be in having this kind of robot as a social companion (all on 0 [not at all] to 10 [completely] scales).

Participants also reported how they would evaluate a company that employed this kind of robot (on 1–7 scales anchored at negative/positive, dislike/like, and bad/good). As a consequential measure, we also told participants that we would be donating $1 on behalf of each participant to an organization working on human-robot relations, and that each participant had to decide which organization we would donate to on their behalf. They were given a choice between the American Society for the Prevention of Cruelty to Robots, which was described as working to advance the development of human-like robots, and the Center for the Study of Existential Risk, which was described as working to prevent the development of human-like robots. Both organizations are real.

We also asked how much the robot seemed to pose a threat to human jobs, human safety, and the distinction between humans and machines in order to compare the effects of mind perception and perceived threats on evaluations of robots. Finally, to ensure that our two videos were as similar as possible, we also asked how engaging and convincing the video was, how knowledgeable the speaker seemed, how concrete the arguments in the video were, and participants' overall mood. All of these measured used 0 (not at all) to 10 (completely) scales.

Results and Discussion

The two videos seemed equally engaging and convincing, and the speaker seemed equally knowledgeable in both videos (t's < 1.32, p's > .194; see Table 1 for full results).

Importantly, the manipulation check showed that the videos successfully altered the belief that robots could have a mind ($M_{mind} = 6.31$, $M_{no\ mind} = 2.06$, $t(94) = 7.45$, $p < .001$). Participants were more comfortable interacting with a robot in each of the three contexts we measured (waitress, sales clerk, and social companion): ANOVA revealed a significant effect of condition, $F(1,293) = 15.91$, $p < .001$, task, $F(1,293) = 30.41$, $p < .001$, and no interaction, $F(1,293) = .07$, $p = .788$. Averaging across the three tasks, participants were more comfortable with the idea of interacting with the robot after learning that robots could have a human-like mind (M = 5.90) compared to learning that robots could have a human-like mind (M = 4.41, $t(94) = 2.85$, $p = .005$; see Table 1 for results broken down by task).

Participants in the pro-robot mind condition also evaluated companies employing robots more positively ($M_{mind} = 4.85$, $M_{no\ mind} = 4.12$, $t(94) = 2.17$, $p = .032$), and were more likely to choose the pro-robot organization for their donation (64.7%) compared to participants who saw the video advocating against the possibility of robot consciousness (44.2%, $\chi^2(1) = 3.19$, $p = .074$). Comfort with the robot in the three contexts mediated the effect of condition on both company evaluations β = -.71, 95% CI [-1.22, -.21] and donation choice β = .12, 95% CI [.04, .24].

	No Mind	Mind	Statistical Test
Engaging	6.64	6.05	$t = 1.05$, $p = .297$
Convincing	6.85	6.72	$t = .23$, $p = .819$
Knowledgeable	7.83	7.19	$t = 1.31$, $p = .194$
Mind (MC)	1.77	6.31	$t = 7.69$, $p < .001$
Waitress	5.20	6.40	$t = 1.90$, $p = .061$
Sales Clerk	5.25	6.79	$t = 2.34$, $p = .018$
Companion	2.65	4.46	$t = 2.82$, $p = .006$
Company Evaluation	4.11	4.87	$t = 2.25$, $p = .027$
Pro-robot Donation	44.2%	64.7%	$\chi^2 = 3.19$, $p = .074$

Table 1. Results of Study 4. All scales were 0–10 except donation choice, which was binary.

The experimental results of Study 4 provide clear evidence for the role of mind perception in shaping consumer reactions to highly human-like robots. Study 4 also provided evidence that firms who employ humanoid robots are evaluated negatively, but that such evaluations are improved if the robots are perceived to have a mind. These results therefore suggest that the makers, marketers, and employers of highly human-like robots can help bring their robots out of the uncanny valley by increasing the perception that they have human-like minds. We will explore practical ways of accomplishing this in Study 6. First, our next study will provide converging evidence for the role of mind perception by comparing two cultures that differ in their perception of minds in inanimate objects.

STUDY 5

In this study, participants were recruited either from the United States or from Japan, two cultures in which the belief that robots have a mind is expected to differ naturally. Religions and philosophies that are common in Japan such as Confucianism, Shinto, and Buddhism teach, for example, that consciousness and animacy live in each and every (animate and inanimate) object; in contrast, Christianity strictly distinguishes objects that do or do not have consciousness (Kitano 2007; Knight 2014). This cross-cultural comparison therefore suggests that Japanese consumers may be more likely to perceive robots as having a human-like mind, which in turn should improve their evaluations of robots.

Method

We showed 150 American MTurk participants (51% female, mean age = 34) and 150 Japanese participants (39% female, mean age = 39) recruited from Lancers, a Japanese equivalent of MTurk a picture of Nadine, a robot with high human-likeness, or a picture of the human being who Nadine was designed to resemble. As in previous studies, participants who saw the picture of the human were told that she was either a human or an advanced humanoid robot. As our primary dependent variable (evaluation of robots), participants then completed the same 3-item measure of uncanniness used in Study 1 (how much would you feel *uneasy/unnerved/creeped out* during an interaction with this robot/person). Each of these items used 0–10 scales.

We also measured cultural constructs that may be relevant to evaluating social robots. Specifically, we measured whether participants believed robots to have a mind (on a binary yes/no scale); how many times they had interacted with a robot in real life (as an open-ended measure of personal familiarity with robots); and finally, how lonely, connected to others, and distant from others they feel (as a measure of loneliness, using 1–5 scales: $\alpha = .83$). We measured familiarity because Japanese society tends to be more familiar with robots than American society (MacDorman, Vasudevan, and Ho 2009). We measured loneliness because Japanese society is often portrayed as suffering from a loneliness epidemic (Onishi 2017), and loneliness has been shown to increase the tendency to anthropomorphize objects (Epley, Waytz, and Cacioppo 2007).

All materials used in this study were translated from English to Japanese by a native Japanese speaker fluent in English, and then back-translated from Japanese to English by a second native Japanese speaker also fluent in English, in order to ensure that the translations captured the original intent of the materials. Japanese participants completed the survey in Japanese, and were required to have been born, raised, and to still live in Japan to be eligible to participate in the study.

Results and Discussion

We found that participants evaluation the robot (i.e., its' uncanniness) was predicted by the robot being evaluated, $F(2,294) = 63.61$, $p < .001$, participants' country, $F(2,294) = 7.31$, $p = .007$ and their interaction, $F(2,294) = 8.27$, $p < .001$ (see Figure 4). Among American participants, responses varied strongly between the robots, $F(1,147) = 52.77$, $p < .001$. The human (portrayed as a human) was the least uncanny (M = 1.69). The perfectly human robot (the human portrayed as a robot) was more uncanny (M = 5.39, $t(92) = 6.87$, $p < .001$ compared to the human), and the Nadine robot was the most uncanny (M = 6.71, $t(97) = 2.17$, $p = .015$ compared to the perfectly human robot).

Responses varied by robot among Japanese participants as well, although the effect was more than four times smaller than it was among Americans, $F(1,147) = 12.53$, $p < .001$. The human-as-human was the least uncanny (M = 2.55). The perfectly human robot was significantly more uncanny (M = 4.26, $t(105) = 3.95$, $p < .001$ compared to the human), and the Nadine robot was again the most uncanny (M = 4.79, $t(105) = 1.20$, $p = .233$ compared to the perfectly human robot). Overall, Japanese participants saw the robots as less uncanny (M = 4.48) than the Americans (M = 6.14, $t(298) = 4.51$, $p < .001$).

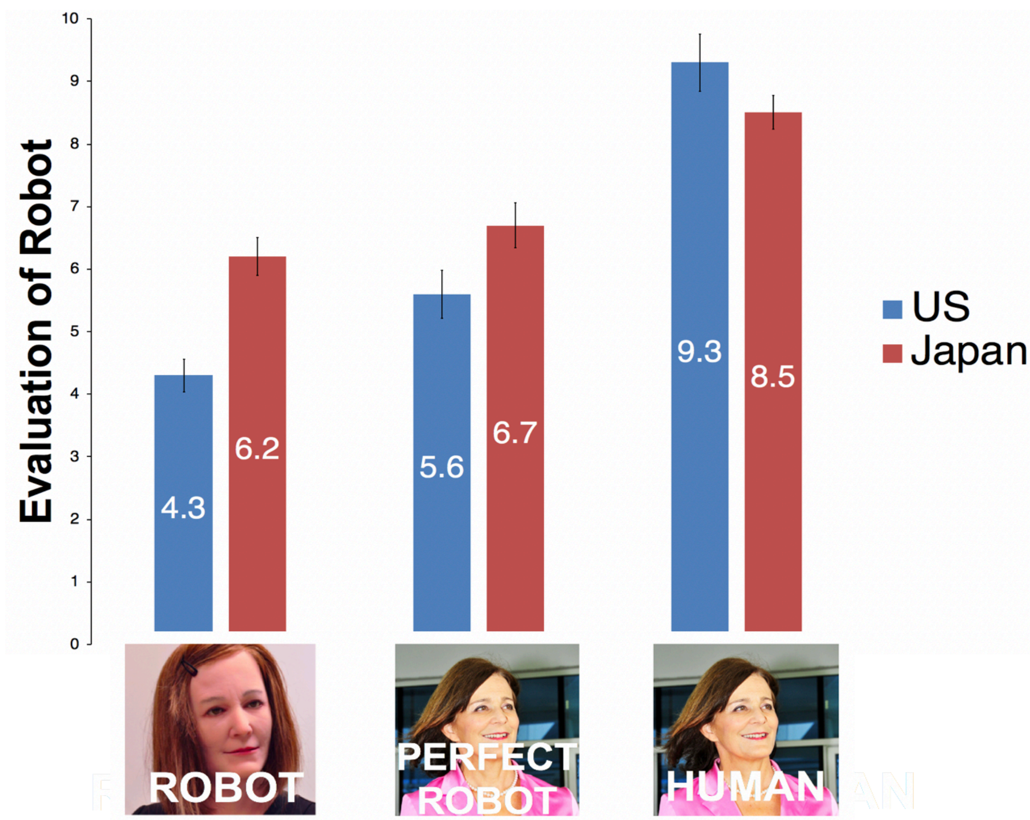


Figure 4. Perceived uncanniness (negative evaluation) to robots and humans among American and Japanese participants. Error bars represent standard errors.

What explains these differences between Japanese and American participants? Japanese participants were more lonely (M = 2.99) than Americans (M = 2.49, $t(298) = 2.83$, $p < .001$) and were more likely to believe that robots could have a mind (21.2% of whom answered yes) than Americans (8.5% of whom answered yes, $\chi^2(1) = 11.67$, $p < .001$). Interestingly, there was no significant difference between the two countries in terms of how many times participants had interacted with a robot in real life ($M_{Japan} = 1.91$, $M_{US} = 2.22$, $t(398) = .53$, $p = .598$). A mediation analysis with country as the independent variable, personal experience with robots, mind perception, and loneliness as the parallel mediators, and uncanniness as the dependent variable revealed that the only significant mediator was mind perception, β = .27, 95% CI [.03–.65]. Neither familiarity with robots, β = -.06, 95% CI [-.41, .25] nor loneliness, β = .22, 95% CI [-.01, .68] mediated the effect of culture on perceived uncanniness of the robots. Thus, Japanese consumers appear to be more likely to perceive robots as having a mind, which in turn improves evaluations of the robots.

The results of Study 5 revealed cross-cultural differences in the evaluation of robots. Despite these cultural differences, however, it is important to note that both groups of participants perceived the more human-like robots (and even the perfectly human-like robot) more negatively than the actual human. Thus, robots with a high (and even perfect) human-likeness elicit relatively negative reactions even in cultures that should be predisposed to seeing robots more positively.

STUDY 6

Our final study has four primary aims. The first is to provide a more direct test of how the two dimensions of mind (agency and experience) contribute to the evaluation of human-like robots. The second is to shed more light on the process by which mind perception improves consumer evaluations of robots, specifically by focusing on the notion of empathy. The third is to address a possible alternative explanation for the out of the valley effect, beyond threats and aesthetics, by showing that mind perception remains a significant contributor to this effect even when controlling for the perceived competence and quality of the robot. Finally, the fourth is to test whether the effects of mind perception might differ depending on the specific task for which the robot is being used.

Research has shown that robots are seen as having moderate levels of agency and very low levels of experience (Gray, Gray, and Wegner 2007), and that negative feelings of uncanniness occur primarily when robots are portrayed as having experience, not agency (Gray and Wegner 2012). Extrapolating from the Gray and Wegener (2012) result would suggest that our finding that mind perception has a positive impact on evaluations of robots should be driven by the agency dimension, because agency (compared to experience) seems more "congruent" with machines. In contrast, the results from Study 2 suggest that both agency and experience have *positive* effects on evaluations.

Indeed, we expect that the positive effects of mind perception in highly human-like robots will be obtained only if the robot is seen as having *both* dimensions of mind. The reason for this proposed divergence from prior research is that Gray and Wegner (2012) used robots with a very low human-likeness as stimuli in their studies, whereas the out of the valley effect occurs with highly human-like robots. We argue that low human-likeness robots are indeed not expected to have the capacity to experience emotion, and that the presence of such capacities would thus be unsettling. In contrast, our findings thus far indicate that the positive effects of mind perception are obtained only with highly human-like robots, for whom we argue the presence of a "complete" mind (agency *and* experience) seems more congruent with their appearance. Indeed, for robots that look almost perfectly human-like, the absence of one dimension of mind might be more unsettling than the presence of a compete mind.

H6: Robots with a complete mind (agency and experience) will be evaluated more positively than robots with only one dimension of mind.

Why does having a mind seem to improve evaluations of human-like robots? As explained in the introduction, we propose that a key reason has to do with empathy. Perceiving a mind in a robot is likely to increase the perception that the robot has something fundamental in common with humans and therefore one may be able to understand as a human what the robot is thinking and feeling. Feeling this sense of commonality and of being able to understand a robot is in turn likely to make consumers feel more comfortable with the robot.

Note that a plausible alternative explanation for the positive effects of mind perception on evaluations of robots might be that informing consumers that a robot has more abilities (i.e., agency *and* experience vs. only one dimension, or having a mind vs. not having a mind) simply makes the robot seem like a more competent, higher quality product. This is related to the notion of aesthetics described earlier – not only can a robot's aesthetic appearance create the impression of ill health, but it can also impact perceptions of product quality and competence. This study will therefore control for perceptions of competence and product quality in order to ensure that mind perception has effects on evaluations over and above these other factors.

Finally, robots can be used to perform many different tasks and the nature of the task may determine whether or not mind perception has a positive effect on evaluations. Specifically, we expect that having a mind will be more beneficial for tasks that are social in nature (i.e., that involve interacting directly with the robot) compared to tasks that are less social and more administrative.

H7: Mind perception will have a larger effect on evaluations of robots when the robot is used for tasks that involve social interaction.

Method

208 participants (51% female, mean age = 34) were recruited from Prolific Academic. Participants were assigned to one of three conditions in a between-subjects design (only autonomy, only experience, or complete mind). All participants were shown a picture of a highly human-like robot (Erica), and were then informed that these kinds of robots could either experience emotions like humans do but could not make autonomous decisions, or could make autonomous decisions like humans do but not experience emotions, or had both capacities. As in Study 2, we chose to focus on emotion specifically (instead of emotion and sensation). When the robots were said to have autonomy and/or experience emotion, we explained that the experience of emotion and the capacity for autonomy is created by a pattern of electrical activity in the human brain, which can be replicated in machines to reproduce the same phenomena in robots. Participants were asked to summarize the information the read about in order to increase engagement with the stimuli.

Participants then completed manipulation checks (i.e., reported how much this kind of robot seems capable of experiencing emotion and making autonomous decisions). As our primary dependent variables, we asked how comfortable participants would feel as a patient in a hospital where such a robot works as (1) a nurse and (2) the hospital administrator. We chose these two jobs in the same setting (i.e., a hospital) because they differ in terms of how much direct social interaction participants could expect to have with the robot. As a measure of empathy as a potential mediator, we asked whether participants thought that their mind and the mind of a robot have anything in common, how much they thought they could understand what the robot is thinking and feeling, and whether they would feel sympathy for the robot if it appeared to be suffering. Finally, to measure a potential alternative explanation based on the perceived competence or quality of the robot, we asked how much the robot seemed competent and how much it seemed like a high-quality product. All measures used 0–10 scales.

Results and Discussion

First, the manipulation checks revealed that our manipulations were effective. The robot seemed to have the greatest capacity for experiencing emotion in the complete mind condition ($M_{\text{complete mind}} = 5.62$), less in the emotion only condition ($M_{\text{emotion}} = 4.54$, $t(134) = 2.13$, $p = .035$), and least in the autonomy only condition ($M_{\text{autonomy}} = 3.78$, $t(140) = 1.51$, $p = .132$ compared to emotion only). The robot seemed to have the most autonomy in the complete mind condition ($M_{\text{complete mind}} = 6.51$), less in the autonomy only condition ($M_{\text{autonomy}} = 5.72$, $t(136) = 1.58$, $p = .115$), and least in the emotion only condition ($M_{\text{emotion}} = 4.86$, $t(140) = 1.76$, $p = .080$ compared to autonomy only).

Turning to the primary dependent variables, ANOVA revealed that condition had a marginally significant effect on participants' comfort with the robot as a nurse, $F(2, 205) = 2.00$, $p = .085$. Planned contrasts showed that participants were most comfortable with the robot when it had a complete mind ($M_{\text{complete mind}} = 3.48$), and less comfortable when it had only one dimension of mind ($M_{\text{emotion}} = 2.50$, $t(134) = 2.10$, $p = .037$; $M_{\text{autonomy}} = 2.69$, $t(136) = 1.67$, $p = .097$). Combining the two single-dimension conditions, the contrast with a complete mind was significant ($M_{\text{single dimension}} = 2.59$, $t(206) = 2.14$, $p = .035$).

For the less social task of hospital administrator, condition had a smaller, non-significant effect on participants' comfort with the robot, $F(2, 205) = 1.13$, $p = .326$. Participants were most comfortable with the robot when it had a complete mind ($M_{complete\ mind} = 3.44$), and about as comfortable when it had only one dimension of mind ($M_{emotion} = 3.69$, $t(134) = 1.50$, $p = .134$; $M_{autonomy} = 2.69$, $t(136) = .83$, $p = .405$). Combining the two single-dimension conditions, the contrast with a complete mind was not significant ($M_{single\ dimension} = 2.85$, $t(206) = 1.33$, $p = .186$). However, the interaction between condition and task (nurse vs. hospital administrator) was not significant, $F(2, 205) = .146$, $p = .864$).

In order to test for the mediating effect of empathy, we conducted a serial mediation analysis in which condition (complete mind vs. single dimension of mind) predicted the perception of mind in the robot (the average of autonomy and emotion; $\beta = 1.34$, $p < .001$, which in turn predicted empathy for the robot (3-item scale, $\alpha = .69$; $\beta = .47$, $p < .001$, which in turn predicted comfort with the robot (average of comfort with robot as nurse and hospital administrator; $\beta = .62$, $p < .001$). The indirect effect was 0.30, 95% CI [.10, .56].

Finally, we confirmed that the effects of mind perception ($\beta = .20$, $p < .001$) and empathy ($\beta = .27$, $p < .001$) on comfort with the robot as a nurse and hospital administrator are robust when controlling for the perceived competence of the robot ($\beta = .31$, $p < .001$) and its perceived quality ($\beta = .01$, $p = .916$).

The results of this study further support mind perception as a robust explanation of the out of the valley effect and lend greater nuance to this explanation by demonstrating that having *both* dimensions of mind leads to more positive evaluations of the robot than having only autonomy or experience alone. Furthermore, this study shows that empathy for the robot partly explain the positive effects of mind perception on evaluations and suggests that mind perception may have a smaller effect in non-social contexts. These results also suggest practical methods of increasing the perceived mind of a robot, by using the emerging technologies of emotional and cognitive computing to endow robots with some semblance of emotional and autonomous capacities respectively. We will discuss these possibilities in more detail in the following General Discussion.

GENERAL DISCUSSION

Social robots have the potential to revolutionize multiple sectors of the economy. They are already being deployed in many ways, from working in retail and service jobs, to assisting the elderly and disabled, to augmenting the productivity of human workers in healthcare, education, and more. The potential for positive impact from these robots depends on humans feeling comfortable around them – and this comfort is far from guaranteed. We showed that robots' human-likeness has initially negative effects on consumer reactions towards these robots, but that these effects reverse past a certain threshold of human-likeness. However, reactions to perfectly human-like robots remain far worse than reactions to humans, even among consumers who believe that robots can have a human-like mind.

The into the valley effect observed in Studies 1 and 2 replicates prior research on the uncanny valley phenomenon and can be explained by theories based on threats and aesthetic factors. However, once robots become very human-like – about the level of human-likeness currently exhibited by the world's most advanced human-like robots – then further increases in human-likeness beyond that point have positive effects on consumer reactions. Yet, even reactions to perfectly human-like robots remain far worse than reactions to real humans. This out of the valley phenomenon has never been demonstrated before, and we have shown that it occurs consistently and can be explained by a theory based on mind perception.

Specifically, we found that there is a positive relationship between a robot's physical human-likeness and its perceived mind, but that perceiving a mind in a robot has positive effects on evaluations only when human-likeness (and therefore perceived mind) is relatively high. We showed that this occurs because perceiving a mind in a robot produces empathy for the robot. Furthermore, we found that these positive effects of mind perception only occur when the robot is perceived to have a complete mind – agency and experience – but not when it has only one of these two dimensions of mind.

Limitations and Future Directions

Given that people would be interacting with robots in real-life personal and business situations, the most important limitation of this research is that participants were not interacting with robots in person, but instead saw pictures or videos of robots interacting with other people. However, this limitation is inevitable in the context of the "out of" the valley phenomenon because the most human-like robots are extremely limited in numbers and extremely costly in monetary and programming terms for conducting research (e.g., including specific experimental manipulations), and of course perfectly human-like robots do not yet exist. Consumer researchers must therefore study this context using the tools currently available given that the emergence of highly human-like robots has important potential implications for consumers and firms. Future research using real interactions between humans and advanced humanoid robots, especially over extended periods of time, would nevertheless be highly valuable, and may be possible in the foreseeable future.

In the present research we did not examine different potential applications of robots in greater detail, instead focusing mainly on consumers' general reactions. In Study 4, we did find that the effects mind perception had the same effects on consumers' comfort with robots as a store clerk, a waitress, and a social companion, suggesting that our effects are likely to be stable across many contexts and applications. Furthermore, in Study 6 we found that the effects of mind perception were similar (although slightly stronger) in a task that involved direct social interaction with the robot (i.e., nurse vs. hospital administrator). Nevertheless, it would be worthwhile to further investigate whether there are specific applications in which human-likeness and mind perception matter more or less to consumers. For example, while hospital administrators do of course interact less directly with patients less than nurses do, there are other jobs in which social interaction plays an even smaller role, and for those tasks we may see that mind perception plays a smaller role.

The finding that an increase in perceived mind affects consumer reactions to robots suggests a practical method of increasing consumers' comfort with robots. The emerging field of affective computing is endowing computers (including robots) with the ability to accurately understand and mimic human emotion, and this ability is being built into products from driverless cars to refrigerators to digital personal assistants (Goasduff 2017; Kodra et al. 2013; McDuff et al. 2013). This technology works by analyzing the facial expressions and tone of voice of the person interacting with the robot. Robots with the ability to understand, express, and react appropriately to humor, frustration, sadness, and other emotions are likely to elicit greater perceptions that they have one of the fundamental dimensions of mind (experience), which in turn is likely to increase consumers' comfort interacting with such robots. Increasing the agency of robots is of course also an ongoing trend in artificial intelligence and robotics research, for example by creating systems that can teach themselves new skills and then exercise those skills independently (Silver et al. 2016), and so future research should also be able to independently manipulate both dimensions of mind in a robot.

In terms of further theory building, future research should explore the lay psychology of mind perception in more detail. Specifically, our results show that consumers' attitudes towards robots improve when they think robots' have a mind, which may suggest that consumers are, philosophically speaking, "lay dualists" whose default belief is that the mind in something immaterial and not feasible for robots to have. Alternatively, consumers may initially be open to a materialist view of the mind and simply be unwilling to grant robots a complete mind for other reasons. One such reason might be what we call "species-ism," meaning that humans may discriminate against robots even if the robots appear to have perfectly human-like appearances and minds, simply because they are not members of the biological human species. Research opportunities may therefore exist in terms of extending research on other forms of discrimination such as sexism and racism into the domain of human-robot interaction. In general, studying the relationships between lay theories of mind and comfort with robots in greater depth seems to be a fruitful path forward.

Conclusion

As consumer robots continue to progress towards greater human-likeness, both in their physical appearance and in their mental abilities, consumers and firms will be faced with unprecedented experiences in stores, restaurants, hotels, hospitals, private homes, and elsewhere. Whereas the majority of consumer robots currently available for purchase are clearly machines for whom questions of mind are irrelevant, the robots of the near future are likely to be so human-like that they will likely be seen as having something like a human-like mind. Increasing perceived mind can improve consumer reactions to these robots – although there may be a hard upper-limit on how much of a mind these robots are granted, which may eventually pose ethical questions about our treatment of robots with experience and autonomy. Indeed, these questions provide the impetus for the organization that we used as a stimulus in our studies: The American Society for the Prevention of Cruelty to Robots, which may one day be taken as seriously as People for the Ethical Treatment of Animals. As robots thus begin to truly blur the line between human and machine, this paper provides guidance for the creators, employers, and marketers of human-like robots to navigate out of the uncanny valley.

APPENDIX – STIMULI USED IN ALL STUDIES

Study 1

Participants watched one of the following videos:

Human as human & human as robot: https://www.youtube.com/watch?v=PM9aG7TIERQ

High human-likeness robot: https://www.youtube.com/watch?v=BrHsDbjkYSU

Low human-likeness robot: https://www.youtube.com/watch?v=H868WZ95Rnw

Participants then answered the following question, all on 0–10 scales:

How much would you feel each of the following emotions during an interaction with this robot [person]?

Uneasy

Unnerved

Creeped out

Study 2

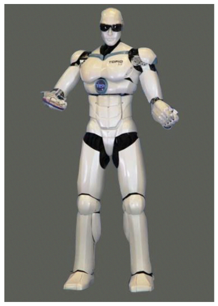
Humans (portrayed either as humans or as robots; participants saw one of four)



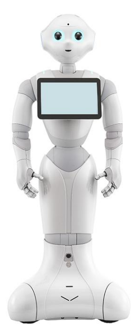
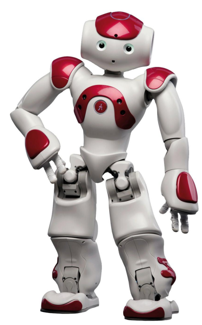
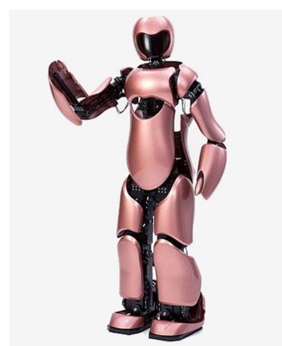

Robots (participants saw one from each of the five following rows)

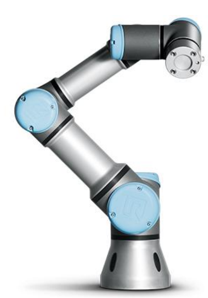




After seeing each robot or human, participants answered the following questions, all on 0–10 scales:

How much do you think this kind of robot could have a mind like humans do?

How much do you think this kind of robot could experience emotion like humans do?

How much do you think this kind of robot could have autonomy like humans do?

How creeped out does this robot [person] make you feel?

How comfortable would you feel interacting with this robot [person]?

How human-like does this robot look overall?

Study 3

Participants in the robot conditions read:

CVS is a chain of pharmacies and drugstores with locations across the United States. CVS is considering a plan to use robots as workers in their stores. These robots would assist customers, answer questions, and work as cashiers. Below is a picture of the kind of robot that they would use.

And saw one of the following pictures:

Participants in the human condition read:

CVS is a chain of pharmacies and drugstores with locations across the United States. CVS is considering hiring the woman below to work in one of their stores.

And saw the picture of the human above.

All participants then answered the following questions, all on 0–10 scales:

How comfortable would you be shopping in a store where this kind of robot [person] is employed?

Do you believe that this robot [person] has a mind of its own?

Do you believe that this robot [person] has a soul?

Do you believe that this robot [person] has consciousness?

I see this kind of robot [person] as a potential threat to human jobs.

I see this kind of robot [person] as a potential threat to human safety.

This kind of robot [person] blurs the line between human beings and machines.

This robot [person] looks like an unhealthy person.

Study 4

Participants watched one of the following videos:

Robots can have a mind: https://www.youtube.com/watch?v=SI4xaihN8Nk

Robots can’t have a mind: https://www.youtube.com/watch?v=QvWQafW3NWg

Participants were then asked:

Please write a few sentences about why the arguments in the video are likely to be true. You can also include any other reasons why you think that robots can eventually [can never] have conscious minds.

Participants were then asked:

Do you believe that robots can eventually have conscious minds?

Participants were then shown the following text and picture:

Several companies are developing robots that look and act almost perfectly like real humans. These robots are being used by businesses as receptionists, salespeople, waiters, and more, and by individual people as social and even romantic companions.

One of these robots, named Sophia, was even granted citizenship by Saudi Arabia earlier this year. The picture below shows one of these advanced humanoid robots. Please answer the following questions about this robot.

Participants then answered the following questions, on 0-10 scales unless specified:

How comfortable would you be shopping in a store where this kind of robot is employed?

How comfortable would you be dining in a restaurant where this kind of robot is employed?

How interested would you be in having this kind of robot as a social companion?

Using the scale below, please indicate the position that best describes your overall evaluation of a company that would employ this kind of robot. [negative/positive, dislike/like, bad/good, on 1-7 scales]

As part of our research on this topic, we will be donating $1 for every participant who completes this survey to an organization that works on human-robot relations. We want to give you the opportunity to decide which organization we donate to on your behalf.

The first organization is called The American Society for the Prevention of Cruelty to Robots. They work to advance the development of human-like robots, which they think will be good for society.

The second organization is called The Center for the Study of Existential Risk. They work to prevent the development of human-like robots, which they think will be bad for society.

Which organization would you like us to donate $1 to on your behalf?

Do you believe that this kind of robot can have consciousness?

Do you believe that this kind of robot can have a mind?

Do you believe that this kind of robot can have a soul?

How much do you think you could understand what this kind robot is thinking and feeling?

I see this kind of robot as a potential threat to human jobs.

I see this kind of robot as a potential threat to human safety.

This kind of robot blurs the line between human beings and machines.

How engaging was the video you saw at the beginning of this survey?

How convincing was the video you saw at the beginning of this survey?

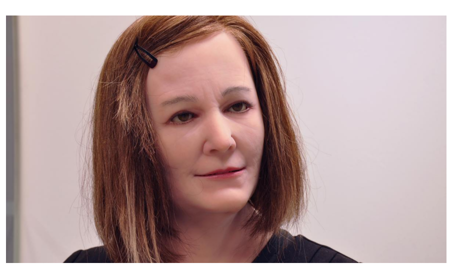
How knowledgeable did the person in the video seem?

Study 5

Participants in the robot conditions read the following:

Several new companies are selling "home robots," such as the Nadine robot below. The robot is meant to be used as a companion and assistant around the house. Nadine can capture photos and videos, play music, monitor the home, look up facts, recognize your voice, respond to commands, and carry on a conversation.

These participants then saw one of the following pictures:



Participants in the human condition read the following and saw the image of the human on the right, above.

Below is a woman named Nadine. We will ask you some questions about her on the next page.

Participants then answered the following questions:

How much would you feel each of the following emotions during an interaction with this robot [person]?

Uneasy

Unnerved

Creeped out

On 0 – 10 scales.

Do you believe that robots have a mind of their own?

Do you believe that robots have a soul?

Do you believe that robots have consciousness?

On Yes/No scales.

How many times have you interacted with a robot in person? [open ended]

How much do you agree with the following statements?

I feel that I am not connected with the people around me.

I often feel that I am alone.

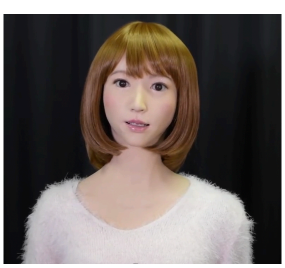
I feel I am distant from the people around me.

On 1 – 5 scales.

Study 6

All participants read the following introduction.

Robots today look and behave more and more like humans. This image shows a robot called Erica, who is currently working as a TV news anchor in Japan. Other similar robots work as receptionists in hotels and department stores, and one robot called Sophia has even been granted citizenship in Saudi Arabia.



Participants then read one of the following:

Experience only:

Beyond *looking* very human-like, these robots also have some components of a conscious mind in the same way that we humans do. Specifically, these robots use a technology called *emotional computing*, which allows them to experience and express real emotions just like we humans do.

This is possible because when humans feel an emotion, the brain produces a pattern of electrical activity than can be replicated in a robot, letting it feel and express emotions just like we do.

However, these robots *cannot* think or act with autonomy like humans do – other than emotion, everything they do and say must be pre-programmed by a human. These robots therefore have some parts of a conscious mind, but not all.

Autonomy only:

Beyond *looking* very human-like, these robots also have some components of a conscious mind in the same way that we humans do. Specifically, these robots use a technology called *cognitive computing*, which allows them to think and act with autonomy, without needing to be pre-programmed by a human.

This is possible because when humans have an independent thought or make an autonomous decision, the brain produces a pattern of electrical activity that can be replicated in a robot, allowing it to plan and act just like we do.

However, these robots *cannot* experience emotions like humans do. These robots therefore have some parts of a conscious mind, but not all.

Complete mind:

Beyond *looking* very human-like, these robots also have the capacity for a conscious mind in the same way that we humans do. Specifically, these robots use a technology called *cognitive computing*, which allows them to think and act with autonomy, without needing to be pre-programmed by a human. When humans have an independent thought or make an autonomous decision, the brain produces electrical and chemical activity that can be replicated in a robot, allowing it to plan and act just like we do.

In addition, these robots use a technology called *emotional computing*, which allows them to experience and express real human emotions. When humans feel an emotion, the brain produces a different pattern of electrical and chemical activity than can be replicated in a robot, allowing it to feel and express emotions just like we do.

These robots therefore have the capacity for all the same components of a conscious mind that we humans do.

Participants then answered the following questions, all on 0–10 scales:

Do you believe that robots will eventually be able to experience emotion like humans can?

Do you believe that robots will eventually have autonomy like humans do?

How comfortable would you feel as a patient in a hospital where this kind of robot is employed as a nurse?

How likely would you be as a patient to visit a hospital where this kind of robot is employed as a nurse?

How comfortable would you feel as a patient in a hospital where this kind of robot is employed as a hospital administrator?

How likely would you be as a patient to visit a hospital where this kind of robot is employed as a hospital administrator?

Do you think that your mind and the mind of a robot have anything in common?

How much do you think you could understand what this kind of robot is thinking and feeling?

Would you feel sympathy for this kind of robot if it seemed to be suffering?

How much do you think this kind of robot could understand what you are thinking and feeling?

This kind of robot seems competent.

This kind of robot seems like a high-quality product.

I see this kind of robot as a potential threat to human jobs.

I see this kind of robot as a potential threat to human safety.

This kind of robot blurs the line between human beings and machines.

REFERENCES

Bartneck, Christoph, Takayuki Kanda, Hiroshi Ishiguro, and Norihiro Hagita (2009), "My Robotic Doppelgänger - A Critical Look at the Uncanny Valley," in *Proceedings - 2009 IEEE International Workshop on Robot and Human Interactive Communication*, 269–76.

Business Insider (2015), "Growth Statistics For Robots Market," *Business Insider*, [available at: http://www.businessinsider.com/growth-statistics-for-robots-market-2015-2].

Business Wire (2017), "$5.43 Billion Consumer Robot Market 2017 - Industry Trends, Opportunities and Forecasts to 2023," *Business Wire*, [available at: https://www.businesswire.com/news/home/20171219005736/en/5.43-Billion-Consumer-Robot-Market-2017].

Dass, Camillia (2017), "Asia's First Retail Robot Launched to Help Relieve Manpower Shortages," *The Straits Times*, [available at: http://www.straitstimes.com/singapore/asias-first-retail-robot-launched-to-help-relieve-manpower-shortages].

Dennett, Daniel (1994), "Consciousness in Human and Robot Minds," in *Cognition Computation and Consciousness*, Oxford: Oxford University Press.

Epley, Nicholas, Adam Waytz, and John T Cacioppo (2007), "On Seeing Human: A Three-Factor Theory of Anthropomorphism," *Psychological Review,* 114(4), 864–86.

Ferrari, Francesco, Maria Paola Paladino, and Jolanda Jetten (2016), "Blurring Human–Machine Distinctions: Anthropomorphic Appearance in Social Robots as a Threat to Human Distinctiveness," *International Journal of Social Robotics*, 8(2), 287–302.

Gibbs, Samuel (2016), "Five in-Home Robots That Could Change Your Life," *The Guardian*, [available at https://www.theguardian.com/technology/shortcuts/2016/jun/01/five-in-home-robots-that-could-change-your-life].

Goasduff, Laurence (2017), "Emotion AI Will Personalize Interactions," *Gartner*, [available at https://www.gartner.com/smarterwithgartner/emotion-ai-will-personalize-interactions/].

Goertzel, Ben (2014), "Artificial General Intelligence: Concept, State of the Art, and Future Prospects," *Journal of Artificial General Intelligence*, 5(1), 1–48.

Goertzel, Ben, David Hanson, and Gino Yu (2014), "A Software Architecture for Generally Intelligent Humanoid Robotics," *Procedia Computer Science*, 41, 158–63.

Gray, Heather, Kurt Gray, and Daniel Wegner (2007), "Dimensions of Mind Perception," *Science*, 619(February), 10–11.

Gray, Kurt and Daniel M. Wegner (2012), "Feeling Robots and Human Zombies: Mind Perception and the Uncanny Valley," *Cognition*, 125(1), 125–30.

Hanson, David (2005), "Expanding the Aesthetic Possibilities for Humanoid Robots," in *IEEE-RAS International Conference on Humanoid Robots*, 24–31.

Higgins, E. Tory and William S. Rholes (1978), "'Saying Is Believing': Effects of Message Modification on Memory and Liking for the Person Described," *Journal of Experimental Social Psychology*, 14(4), 363–78.

Ishiguro, Hiroshi and Shuichi Nishio (2007), "Building Artificial Humans to Understand Humans," *Journal of Artificial Organs*, 10(3), 133–42.

Jackson, Joseph (1993), "Realistic Group Conflict Theory: A Review and Evaluation of the Theoretical and Empirical Literature," *The Psychological Record*, 43(3), 395–414.

Kätsyri, Jari, Klaus Förger, Meeri Mäkäräinen, and Tapio Takala (2015), "A Review of Empirical Evidence on Different Uncanny Valley Hypotheses: Support for Perceptual Mismatch as One Road to the Valley of Eeriness," *Frontiers in Psychology*, 6, 390.

Kitano, Naho (2007), "Animism, Rinri, Modernization; the Base of Japanese Robotics," *ICRA 2007 Roboethics Workshop*, 1–4.

Knight, Heather (2014), *How Humans Respond to Robots: Building Public Policy through Good Design*, [available at http://www.brookings.edu/~/media/Research/Files/Reports/2014/07/29 how humans respond to robots knight/HumanRobot PartnershipsR2.pdf].

Kodra, Evan, Thibaud Senechal, Daniel McDuff, and Rana El Kaliouby (2013), "From Dials to Facial Coding: Automated Detection of Spontaneous Facial Expressions for Media Research," in *2013 10th IEEE International Conference and Workshops on Automatic Face and Gesture Recognition, FG 2013*.

MacDorman, Karl F., Sandosh K. Vasudevan, and Chin Chang Ho (2009), "Does Japan Really Have Robot Mania? Comparing Attitudes by Implicit and Explicit Measures," *AI and Society*, 23(4), 485–510.

Mathur, Maya B. and David B. Reichling (2016), "Navigating a Social World with Robot Partners: A Quantitative Cartography of the Uncanny Valley," *Cognition*, 146, 22–32.

McDuff, Daniel, Rana El Kaliouby, Thibaud Senechal, May Amr, Jeffrey F. Cohn, and Rosalind Picard (2013), "Affectiva-Mit Facial Expression Dataset (AM-FED): Naturalistic and Spontaneous Facial Expressions Collected 'in-the-Wild,'" in *IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops*, 881–88.

Medin, Douglas and Edward Smith (1984), "Concepts and Concept Formation," *Annual Review* *of Psychology*, 35, 113–38.

Mitchell, Wade J., Kevin A. Szerszen, Amy Shirong Lu, Paul W. Schermerhorn, Matthias Scheutz, and Karl F. MacDorman (2011), "A Mismatch in the Human Realism of Face and Voice Produces an Uncanny Valley," *i-Perception*, 2(1), 10–12.

Mori, Masahiro (1970), "The Uncanny Valley," *Energy*, 7(4), 33–35.

Müller, Vincent C. and Nick Bostrom (2016), "Future Progress in Artificial Intelligence: A Survey of Expert Opinion," in *Fundamental Issues of Artificial Intelligence*, Cham: Springer, 555–72.

Nguyen, Clinton (2016), "Chinese Restaurants Are Replacing Waiters with Robots," *Business Insider*, [available at http://www.businessinsider.com/chinese-restaurant-robot-waiters-2016-7/].

Onishi, Norimitsu (2017), "A Generation in Japan Faces a Lonely Death," *New York Times*, [available at https://www.nytimes.com/2017/11/30/world/asia/japan-lonely-deaths-the-end.html].

Peer, Eyal, Laura Brandimarte, Sonam Samat, and Alessandro Acquisti (2017), "Beyond the Turk: Alternative Platforms for Crowdsourcing Behavioral Research," *Journal of Experimental Social Psychology*, 70, 153–63

PwC (2017), "Will Robots Steal Our Jobs? The Potential Impact of Automation," *UK Economic Outlook*, [available at https://www.pwc.co.uk/economic-services/ukeo/pwcukeo-section-4-automation-march-2017-v2.pdf].

Riek, Blake M., Eric W. Mania, and Samuel L. Gaertner (2006), "Intergroup Threat and Outgroup Attitudes: A Meta-Analytic Review," *Personality and Social Psychology Review*, 10(4), 336–53.

Rory Cellan-Jones (2014), “Stephen Hawking Warns Artificial Intelligence Could End Mankind,” *BBC News*, [available at http://www.bbc.com/news/technology-30290540].

Rosenthal-Von Der Pütten, Astrid M. and Nicole C. Krämer (2014), “How Design Characteristics of Robots Determine Evaluation and Uncanny Valley Related Responses,” *Computers in Human Behavior*, 36, 422–39.

Sandberg, Anders and Nick Bostrom (2008), *Global Catastrophic Risks Survey*, [avialble at https://www.fhi.ox.ac.uk/reports/2008-1.pdf].

Searle, John Rogers (1995), *The Rediscovery of the Mind*. Cambridge: MIT Press.

Silver, David, Aja Huang, Chris J Maddison, Arthur Guez, Laurent Sifre, George van den Driessche, Julian Schrittwieser, Ioannis Antonoglou, Veda Panneershelvam, Marc Lanctot, Sander Dieleman, Dominik Grewe, John Nham, Nal Kalchbrenner, Ilya Sutskever, Timothy Lillicrap, Madeleine Leach, Koray Kavukcuoglu, Thore Graepel, and Demis Hassabis (2016), “Mastering the Game of Go with Deep Neural Networks and Tree Search.,” *Nature*, 529(7587), 484–89.

Simon, Matt (2015), “This Incredible Hospital Robot Is Saving Lives. Also, I Hate It,” *WIRED*, [available at https://www.wired.com/2015/02/incredible-hospital-robot-saving-lives-also-hate/].

Simonsohn, Uri (2017), “Two-Lines: A Valid Alternative to the Invalid Testing of U-Shaped Relationships with Quadratic Regressions,” *University of Pennsylvannia working paper.*

Wang, Shensheng, Scott O Lilienfeld, and Philippe Rochat (2015), “The Uncanny Valley: Existence and Explanations,” *Review of General Psychology*, 19(4), 393–407.

Waytz, Adam, Kurt Gray, Nicholas Epley, and Daniel M. Wegner (2010), "Causes and Consequences of Mind Perception," *Trends in Cognitive Sciences*, 383–88.

Zlotowski, Jakub, Diane Proudfoot, and Christoph Bartneck (2013), "More Human Than Human: Does The Uncanny Curve Really Matter?," *Proceedings of the HRI 2013 Workshop on Design of Humanlikeness in HRI from uncanny valley to minimal design*, (January), 7–13.