



Article

First encounter with robot Alpha: How individual differences interact with vocal and kinetic cues in users' social responses

new media & society

1–26

© The Author(s) 2019

Article reuse guidelines:

sagepub.com/journals-permissions

DOI: [10.1177/1461444819851479](https://doi.org/10.1177/1461444819851479)

journals.sagepub.com/home/nms



Kun Xu 

University of Florida, USA

Abstract

The Computers are Social Actors (CASA) paradigm was proposed more than two decades ago to understand humans' interaction with computer technologies. Today, as emerging technologies like social robots become more personal and persuasive, questions of how users respond to them socially, what individual factors leverage the relationship, and what constitutes the social influence of these technologies need to be addressed. A lab experiment was conducted to examine the interactions between individual differences and social robots' vocal and kinetic cues. Results suggested that users developed more trust in a social robot with a human voice than with a synthetic voice. Users also developed more intimacy and interest in the social robot when it was paired with humanlike gestures. Moreover, individual differences including users' gender, attitudes toward robots, and robot exposure affected their psychological responses. The theoretical, practical, and ethical value of the findings was further discussed in the study.

Keywords

Artificial intelligence, gestures, human–robot interaction, social cues, social presence, social robots, the Computers are Social Actors paradigm

Introduction

The past few years have witnessed an increasing number of popular media portrayals that feature humans' encounter with social robots. Television shows such as *Westworld*

Corresponding author:

Kun Xu, Department of Telecommunication, College of Journalism and Communications, University of Florida, Gainesville, FL 32611, USA.

Email: bigben219@gmail.com

and *Humans* have aroused much discussion about the distinction between humans and robots. While these fictional portrayals of the human–robot relationship may still be distant from our real life human–robot interaction (HRI), scholars have seen a growing trend to apply social robots in family communication, health communication, and education. For instance, Robot Kirobo was among the first social robots to be sent to the International Space Station as a companion of a Japanese astronaut (Gannon, 2013). Telepresence robots, defined as devices that incorporate video conferencing and robotic remote control by human operators (Herring, 2004), have been used to connect homebound students to real classroom environments. Robot SAM is designed to take care of elderly people and enhance the communication efficiency between patients and doctors (Bartneck et al., 2008). Robot NAO can help autistic children focus their attention and express their emotions (Duquett et al., 2008). Given the growing use of these social robots in our society, understanding the applications and the effects of social robots will guide our future interactions with these emerging technologies and further inform technology innovation and diffusion.

One approach to studying humans' social interaction with social robots is to look into the Computers are Social Actors (CASA) paradigm (Nass et al., 1994) and examine how users perceive and respond to social robots as social entities. While the CASA paradigm has demonstrated that users apply social scripts in human–computer interaction, Nass and Moon (2000) proposed a list of questions that future research should investigate. One set of questions asked whether some social dimensions of a computer are more powerful than others and how these dimensions exert synergetic effects on users' social responses. Another set of questions called for research on whether users' social responses are confined to any individual differences and contextual factors. Therefore, following these two general research questions, this study focuses on users' first encounter with a social robot and evaluates the explanatory power of the CASA paradigm in HRI, the influence of the vocal and kinetic dimensions on users' reactions to the social robot, and the role of individual factors including gender, robot exposure, and attitudes toward robots in their psychological processing of the robot.

Examining these questions will yield both theoretical and practical value. In addition to recognizing whether social robots will elicit users' social perception and social attitudes, the study will reflect the different power of each single social cues. That is, instead of gauging the quantitative value of the social cues (Breazeal, 2003; Gong, 2008), the study will explore the quality of social cues and help establish a hierarchy of social dimensions. The findings will render the CASA paradigm more testable and contribute to the theory construction of the paradigm in future human-machine communication research.

Practically, understanding the power of each social dimensions can help customize the interface of social robots. Product designers can use the knowledge to prioritize some cues over others to augment the perceived socialness of the robots and control the budget in refining these emerging technologies. Researchers can also design better user experiences based on individual preferences (Jung and Lee, 2004). Embedding these social cues can help facilitate the collaboration and enhance the communication effectiveness between humans and machines.

Literature review

The CASA paradigm in human–robot interaction

While past works have explored the human–computer relationships from the perspectives of persuasion, ubiquity, affection, mental modeling, and interface design (Engelbart, 1962; Fogg, 2002; Licklider and Taylor, 1968; Norman, 1988; Picard, 1995; Weiser, 1991; Wiener, 1950), the field of human–computer interaction (HCI) can be traced to the classic thought experiment “the imitation game,” where Turing (1950) proposed the question of whether machines can think. Following the discussion of humans’ differences from machines, scholars like Searle (1980) and Suchman (2007) pointed out that machine intelligence was merely a reflection of codes and programs. The debate about human intelligence and machine intelligence has spurred succeeding scholars to fathom the boundaries between humans and machines and the ways humans should interpret and interact with these technologies.

To understand how individuals perceive computers as humans, Nass et al. (1994) proposed the CASA paradigm and found that users’ responses to computers are fundamentally social and natural. Reeves and Nass (1996) later published the book *The Media Equation*, which refers to the idea that humans treat media like real people. Based on the theoretical framework, Nass and colleagues found that even though computer users were aware of the nature of the machines, they exhibited etiquette to computers (Nass, 2004), formed team relations with computers (Nass et al., 1996), perceived computers to have personalities (Nass and Lee, 2001), and perceived computers to have gender differences (Nass et al., 1997). More recently, researchers have applied the CASA paradigm to a range of technologies including computer agents and virtual assistants (Edwards et al., 2014; Guzman, 2019; Spence et al., 2014). For instance, Jung et al. (2014) found that an embedded tutor agent was perceived as more person-like and more attractive than an external agent designed in a separate tutoring device.

While a growing number of studies have centered on users’ interaction with social robots, scholars have conceptualized social robots differently. Given the emphases on social norms, Bartneck and Forlizzi (2004) defined a social robot as “an autonomous or semi-autonomous robot that interacts and communicates with humans by following the behavioral norms expected by the people with whom the robot is intended to interact” (p. 592). Comparatively, Duffy (2003) paid more attention to robots’ materiality and referred to a social robot as “a physical entity embodied in a complex, dynamic, and social environment sufficiently empowered to behave in a manner conducive to its own goals and those of its community” (p. 177). Zhao (2006) took a sociological perspective and defined them as “human-made autonomous entities that interact with humans in a humanlike way” (p. 405). Scholars have also underlined the communication process between humans and robots. For instance, Lee and colleagues suggested that the primary function of social robots is to interact with humans (Lee et al., 2006). While scholars may diverge in whether social robots should be embodied, humanlike, or fully autonomous (Breazeal, 2003; Li, 2015; Pfeifer and Scheier, 1999; Zhao, 2006), social robots should at least feature a certain degree of automation and be partly used for social interaction with humans.

Much HRI research has focused on the transmission of social signals between humans and robots (Lee et al., 2005, 2006; Li et al., 2015). For instance, one recent study has suggested that participants experienced stronger physiological arousal when touching the social robot NAO's sensitive parts like buttocks (Li et al., 2017). Jung and Lee (2004) investigated the influence of physical embodiment and found that a physically embodied zoomorphic social robot was perceived as more attractive than an animation-based one. These studies have indicated that the social cues of social robots can activate users' social reactions.

Despite the association between social cues and social responses, theories have diverged in the effects of social cues on social responses. Following prior research on the quantity of social cues and the discussion of whether more social cues will be more likely to elicit social responses (Kim et al., 2013; Tung and Deng, 2007), scholars have begun to notice the quality of social cues and postulated that the more humanlike the cues are, the stronger social responses users will demonstrate. For example, Hinds et al. (2004) found that participants felt less responsible for a task when collaborating with a humanlike robot than with a machinelike robot partner. Sah and Peng (2015) noticed that a health website with anthropomorphic language including active voice and personal nouns evoked more self-disclosure than the same website without these cues. More recently, scholars have attempted to systematically classify the distinct influence of social cues on users' social responses to media technologies. Xu and Lombard (2016) classified social cues into primary ones and secondary ones. While primary cues are salient and essential to users' interpretation of socialness, secondary social cues are neither sufficient nor necessary in leading to users' social attitudes and behavior. Primary social cues such as human voice, human shape, and eye contact should be more likely to evoke users' social responses than secondary social cues such as human size, text, and movements.

While the majority of past works has highlighted a positive relationship between the human features of social cues and the strength of users' social responses, another thread of research posits that machinelike cues may overtake humanlike cues in activating users' positive attitudes toward machines. Sundar (2008) mentioned that if a technology appears machinelike, users will perceive it as objective, credible, and fair. Alternatively, if an interface affords an anthropomorphic social actor, users may be detracted from the machine heuristic and attribute less objectivity to the interface. As an example, Sundar and Nass (2001) found that participants evaluated the news stories selected by computerized technologies to be in higher quality than those selected by human editors.

While these two perspectives of users' psychological responses to machines may seem paradoxical, their application could be contingent upon specific communication contexts. This study selects a socioemotional context in HRI and compares the effects of two pairs of social cues: human voice versus synthetic voice and gestural movements versus non-gestural movements. As technology users tend to seek more social cues in socioemotional contexts than in task-oriented contexts (Derks et al., 2007; Gunawardena and Zittle, 1997), this study postulates that cues with more humanlike characteristics will elicit stronger social responses.

Vocal and kinetic cues

Social cues in HRI studies have been referred to as "biologically and physically determined features salient to observers because of their potential as channels of useful

information” (Fiore et al., 2013: 2). Examples of these cues include eye contact, blinking, leaning forward, and speech. Among these social cues, this study specifically examines the combination of vocal and kinetic cues on social robots, as prior research has shown that movements and speech together can improve people’s learning efficiency, boost their recall performances, and generate more meaningful social interactions (Cabibihan et al., 2012; Salem et al., 2010, 2011; Sirkin and Ju, 2012). In addition, as Nass and Moon (2000) called for research on “additive or synergetic effects” of different combinations of social dimensions in HCI (p. 98), this study explores both the main effects and the interaction effects of a robot’s vocal and kinetic cues.

Vocal cues. Human voice is defined as “the sound produced by humans and other vertebrates using the lungs and the vocal folds in the larynx or voice box” (National Institute of Health (NIH), 2017). It carries a speaker’s tones, emotions, attitudes, and even social identities. Nass and Brave (2005) argued that when human voice is embedded in computers, it is perceived as a natural and powerful modality in HCI. Comparatively, synthetic voice is sometimes perceived as unnatural and unpleasant (Gong and Lai, 2003). Although synthetic voice can communicate the same content of the messages as human voice, it reduces the effects of paralinguistic cues such as tones and accents (Nass and Lee, 2001). But meanwhile, compared with the high cost of standardizing human-sounding speech, transferring text to synthetic speech is more affordable and less time-consuming (Gong and Lai, 2003).

Both human voice and synthetic voice have been examined in HCI and computer-mediated communication (CMC) contexts (Mayer et al., 2003; Tsimhoni et al., 2001; Wang et al., 2007); however, prior research has rendered inconsistent findings. Nass and Steuer (1993) found that participants were more sensitive to human voice and perceived different voices rather than different computers to be distinct information sources. Bracken and Lombard (2004) also found that children socially responded to a computer’s human-voiced praise. However, Gong and Lai’s (2003) research suggested that a user interface with only a synthetic speech was perceived as more articulate and pleasant than that with a mixed speech that combined both human voice and synthetic voice. Gong and Nass (2007) further noted that participants’ social reactions were undermined when a digital social actor was paired with a human voice and a computerized face. These studies imply that the relationship between human voice and users’ psychological responses may not be linear and straightforward. Considering that scholars have not reached consensus on the social influence of human voice versus synthetic voice, this study applies the CASA paradigm and tests whether human voice is more likely than synthetic voice to evoke users’ social responses to social robots.

Kinetic cues. Humans are evolved to be sensitive to the motion of objects. The Heider and Simmel (1944) experiment illustrated how humans can naturally attribute behavioral intentions to moving dots. Johansson’s (1973) research further demonstrated that when humans are exposed to multiple moving light dots which represent human body joints, they can rapidly identify these dots as a walking person. Ju and Takayama (2009) manipulated the trajectories and the speed of automatic door movements and noted that humans could even perceive door movements to have human cognition and motivations.

While these studies have endorsed the social influence of simple and random movements, social robots are often designed with gestures to augment their engagement in communication (Salem et al., 2010). Gestures are conceptualized as humanlike movements that deliver communicators' purposes, capabilities, mental states, and social rituals (Cabibihan et al., 2012; Hoffman and Ju, 2012). For instance, closed arm positions could mean rejection (Machotka, 1965). Reclining body angle or decreasing backward lean of torso could denote affinity with communication partners (Mehrabian, 1969).

The differences between movements and gestures have been further explicated in Krauss et al.'s (1996) study, where movements were categorized into adapters, symbolic gestures, and conversational gestures. Adapters are mere hand movements that are irrelevant to communicators' intentions or speech content. Symbolic gestures have specific conventionalized meanings and can deliver meanings without auxiliary speech (e.g. thumbs-up and clenching one's fist). Conversational gestures are the concomitants of speech, which can further be sorted into two types: motor gestures and lexical gestures. Motor gestures are intuitive and automatic movements that afford no semantic implications of the accompanying speech (e.g. slightly rolling hands in front of chest to express oneself). Lexical gestures are the movements that reflect the linguistic messages of the speech (e.g. showing directions or describing the size of an object).

Some previous HRI research has shown that users can respond to social robots' gestures in a social manner. Bevan and Fraser (2015) found that those who shook hands with a social robot before negotiating with it were more likely to reach an agreement. Fiore et al. (2013) found that participants reported the robot to be more socially present, friendlier, and politer when a walking social robot got in the way of participants' travel paths and then yielded to the participants. While these experimental studies have revealed the effects of robots with gestures and ones without any movements, limited research has noted the conflation between gestures and movements and explicated whether it is the movements per se or the socially constructed gestures that sway users' mental and behavioral states. Even if gestural movements may deliver more social and cultural meanings than non-gestural movements, the machine heuristics may still lead users to perceive non-gestural kinetic cues as more direct, approachable, and credible (Ju and Sirkin, 2010). Hence, this study seeks to distinguish and evaluate the effects of robots' gestural versus non-gestural movements.

Social presence and social attitudes

Social presence and social attitudes have been studied as indicators of users' social responses. Lee (2004) defined social presence as "a psychological state in which virtual (para-authentic or artificial) social actors are experienced as actual social actors in either sensory or non-sensory ways" (p. 44). Biocca et al. (2003) and Zhao (2003) conceptualized it as the sense of being with others. As social presence may occur in both CMC and HCI contexts, Lombard and Ditton (1997) differentiated social-actor-within-medium presence and medium-as-social-actor presence. Medium-as-social-actor presence occurs when users respond to the social cues presented by the media technologies per se. As this study examines users' social presence experiences in communication with robots, the term "medium-as-social-actor presence" is adopted for clarity and precision.

Social attitudes toward technologies include the extent to which users feel attracted to media technologies, trust in them, and intend to use them in the future (Jung et al., 2014; Nass et al., 1996; Shin and Choo, 2011). Prior research has found positive influence of social cues on social attraction (Jung et al., 2014; Jung and Lee, 2004; Nass and Lee, 2001). For example, researchers found that uniform group identity cues would increase the social attraction among group members (Carr et al., 2011; Postmes et al., 1998). The power of social cues on perceived trustworthiness of technologies has also been corroborated in previous research where bandwagon cues would shape users' evaluation of online products (Sundar et al., 2009). Stoll et al. (2016) indicated that a telepresence robot Double was perceived as more credible when it gave up using any guilt-involving conversation skills in negotiation. In addition, Salem et al. (2013) found that participants reported greater intention of future use when a robot occasionally made mistakes, which highlights users' preference for an imperfect robot. Even minimal group paradigm can be applied to HRI where participants were more willing to accept an in-group robot than an out-group robot (Eyssel and Kuchenbrandt, 2012). Based on previous works, this study evaluates the links between social cues and users' social responses and applies the CASA paradigm to explain users' first encounter with a social robot. The following hypotheses and research question are proposed:

H1. Compared with one with a synthetic voice, a social robot with a human voice will lead to greater levels of (a) medium-as-social-actor presence, (b) perceived attraction of the robot, (c) perceived trustworthiness of the robot, and (d) intention of future use.

H2. Compared with one with non-gestural movements, a social robot with gestures will lead to greater levels of (a) medium-as-social-actor presence, (b) perceived attraction of the robot, (c) perceived trustworthiness of the robot, and (d) intention of future use.

RQ1. How will a social robot's vocal cues interact with its kinetic cues in predicting (a) medium-as-social-actor presence, (b) perceived attraction of the robot, (c) perceived trustworthiness of the robot, and (d) intention of future use?

Individual differences

Oh et al.'s (2018) meta-analysis on the antecedents of social presence indicated that among technology features, contextual factors, and individual differences, least research has explored individual differences in users' social presence experiences. Thus, this study seeks to add to the literature on individual differences in HRI and examines the roles of users' gender, robot use experiences, and attitudes toward robots in their psychological processing of social robots.

Although gender is one of the most commonly examined individual factors in relation to social presence, scholars have not reached consensus on the gender effects (Oh et al., 2018). Nass et al. (1995) found that although males were more likely than females to accept computers taking roles such as babysitters and judges, gender was not related to users' psychological anthropomorphism of the computers. Lee (2008) found that females

were likely to show positive responses to a flattering computer with a human voice; however, when a computer was assigned a machine voice, gender did not make a difference in users' responses.

Technology use experiences may affect users' expectations for the technology performances. Whereas Nass et al. (1995) did not find a significant relationship between computer use experiences and users' attitudes toward computers, Johnson et al. (2004) suggested that only experienced computer users were vulnerable to flattery from computers. Due to the inconsistency of the relationship between users' expertise in computers and their social attitudes toward computers, this study further investigates how users' robot use experiences affect their communication with robots.

The degree to which users accept robots with social roles may also determine their attachment for social robots. A study in Japan showed that while adults expected robots to engage in household duties in the future, elderly people expected robots to serve more in public settings (Nomura et al., 2009). Copleston and Bugmann (2008) found that house cleaning, preparing tea, and washing up appeared to be the top chores that users expected robots to do, but users were not comfortable with robots taking the jobs related to family help, pet care, or security. Thus, this study further explores how users' attitudes toward robots affect their communication with robots. We propose the following research questions:

RQ2. How will users' gender interact with the social cues in predicting users' social responses?

RQ3. How will users' robot use experiences interact with the social cues in predicting users' social responses?

RQ4. How will users' attitudes toward robots interact with the social cues in predicting users' social responses?

Method

Participants

A total of 110 students from a public university in the Northeastern United States voluntarily participated in an experiment. All participants received extra credit for their participation. Those who had experiences in programming robotic technologies or had seen/used the robot adopted in the experiment were excluded. Among the 110 participants, 55 of them were males (50.0%) and the others were females (50.0%; $M=1.50$, $SD=0.50$). They were from 18 years to 34 years old ($M=20.44$, $SD=3.41$).

Research design and procedures

The experiment used a 2×2 between-subjects factorial design. Participants were randomly assigned to one of the four conditions: human voice with gestural movements, synthetic voice with gestural movements, human voice with non-gestural movements, or synthetic voice with non-gestural movements. A social robot "Alpha" (UBTECH, 2017)

was used as the experiment apparatus. Alpha is 15.67 inches tall, 8.19 inches wide, and 4.80 inches deep. There are 16 servomotor joints on its body and limbs. The robot has a mono aural amplifier and can thus speak to users.

To create the human voice, pre-recorded human-voice-based messages were installed in the robot. A male voice rather than a female voice was used because the shape of the robot was designed to have male characteristics. Also, the original speech demo uses a male voice. To create the synthetic voice, the text-to-speech software “SayIt” was used. The same messages were uploaded to the software and transformed into a synthetic voice. After testing different versions of the synthetic voices, the “Bruce” voice was used.

To create the gestural movements of the robot, symbolic gestures and conversational gestures (Krauss et al., 1996) were programmed into the robot Alpha. A list of these gestures was identified in prior research (Cabibihan et al., 2012; Hoffman and Ju, 2012; Salem et al., 2013). Examples of these gestures include opening arms (i.e. to show sincerity and openness), both hands in front of the chest (i.e. to introduce itself), waving one hand (i.e. greetings or saying goodbye), bowing (i.e. to greet or express appreciation), and waving both arms (i.e. cheering). By contrast, random movements that were designed not to communicate specific meanings or relate to the linguistic messages were programmed as the non-gestural movements. For example, the robot was programmed to hold hands flat, raise hands up, and stretch both arms forward or backward during the speech. The number of the movements and the timing of displaying these movements in the non-gestural movement conditions was manipulated to be the same as the ones in the gesture conditions. Figure 1 shows the comparison between a gestural movement and a non-gestural movement. Manipulation checks were conducted to make sure that the research design was successful.

After the participants entered the lab and read the consent form on a laptop, they were asked to answer questions about their demographic information via the survey software Qualtrics. Then, they were led to another table where the robot Alpha was standing. The participants were told that the social robot would give a 2-minute self-introduction. Only one participant participated in the experiment at a time.

The robot Alpha was put about 25 inches away from the participants. The participants were not allowed to touch the social robot. To control for the effects of the distance between the participants and the robot, they were also advised not to change their seat positions.

After the participants were ready, the social robot began its self-introduction which includes four parts. In the first part, the robot introduced its name, where it was made, and its basic functions. In the second part, the robot introduced its potential applications. In the third part, it briefly introduced its experiences of interacting with humans and its feelings for humans’ achievements. Last, it expressed its vision for the future and said goodbye to the participants. The robot made movements and speech simultaneously. Self-referential statements were used in the robot’s self-introduction as most social robots in the market are designed to use subjective personal nouns. Using these self-referential statements increases the external validity of the study. Finally, participants were asked to fill out the questionnaire items for dependent variables and other control variables.

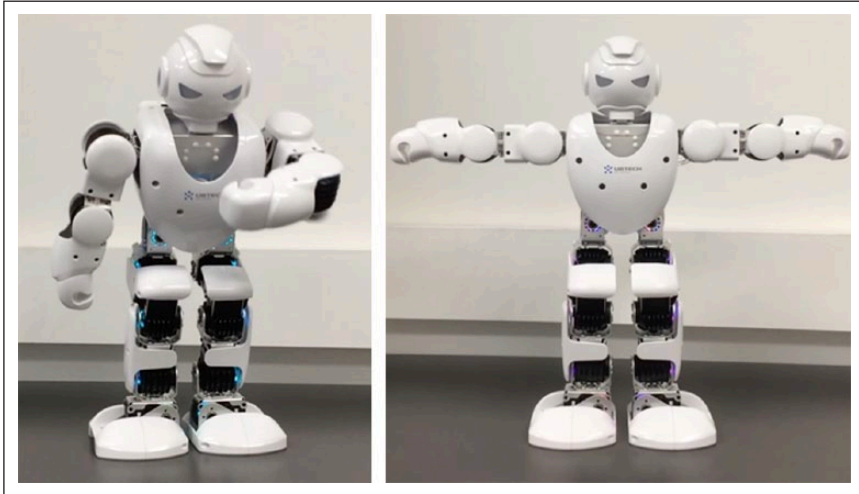


Figure 1. Robot Alpha bows to participants/holds arms flat.

Measures

The measure of medium-as-social-actor presence ($M=6.83$, $SD=1.79$, $\alpha=.87$) was adapted from the measures of social presence in the contexts of HCI and HRI (Lee et al., 2006; Nass and Lee, 2001). Participants were asked to report on a Likert-type scale with seven 10-point statements (1 = not at all, 10 = very much). The statements include “How much did you feel as if you were interacting with an intelligent being?” “How much did you feel as if you and the robot Alpha were communicating with each other?” “How much did you feel as if you were together with an intelligent being?” “How much did you feel as if you were alone (reverse coding)?” “How much attention did you pay to the robot Alpha?” “How much did you feel involved with the robot Alpha?” “How much did you feel as if the robot Alpha was talking to you?”

The measure of perceived social attraction ($M=5.65$, $SD=2.51$, $\alpha=.90$) was adapted from previous measures of social attraction (Lee et al., 2006; McCroskey and McCain, 1974). Participants were asked to report on a 10-point Likert-type scale with four statements (1 = strongly disagree, 10 = strongly agree). Examples of the statements include “I think I could establish a personal relationship with the robot Alpha” and “I think I could have a good time with the robot Alpha.”

The measure of perceived trustworthiness ($M=7.96$, $SD=1.68$, $\alpha=.85$) was adapted from previous measures of trust (Gong, 2008; Gong and Nass, 2007). Participants were asked how they felt about the robot Alpha on a 10-point semantic differential scale with four items. Examples of these items include “untrustworthy–trustworthy” and “unreliable–reliable.”

The measure of intention of future use ($M=7.73$, $SD=2.28$, $\alpha=.92$) was adapted from Shin and Choo’s (2011) and Eyssel and Kuchenbrandt’s (2012) measures of intention to use. The measure used a 10-point Likert-type scale with three statements (1 = strongly

disagree, 10=strongly agree). Examples of the statements include “I would like to use a robot like Alpha again” and “I recommend others to use a robot like Alpha.”

Robot use experiences was adapted from previous measures of computer use experiences (Johnson et al., 2004; Nass et al., 1995). Participants were asked how many times they interacted with a humanoid social robot in the previous year. Zhao’s (2006) definition of a humanoid social robot was provided as a reference in the questionnaire. Results were recoded into a binary variable (0=never, 1=at least once; $M=0.57$, $SD=0.50$).

Participants’ attitudes toward robots taking social roles were adapted from Nass et al.’s (1995) measure on users’ attitudes toward computers with routinized roles. Participants were asked to report on a 6-point scale (1=very uncomfortable, 6=very comfortable). The item is “How comfortable would you be with robots taking routinized roles (e.g. accountants, bank tellers)” ($M=3.55$, $SD=1.70$).

The measure of physical anthropomorphism was adapted from previous measures of anthropomorphism (Kim and Sundar, 2012; Powers and Kiesler, 2006) and used for manipulation checks. Participants were asked to indicate their perception of the robot’s voice ($M=4.49$, $SD=2.64$, $\alpha=.94$) and movements ($M=3.83$, $SD=2.08$, $\alpha=.89$). The measure used a 10-point semantic differential scale with three items: “machinelike–humanlike,” “unnatural–natural,” and “artificial–lifelike.” Participants were also asked whether and why they perceived the robot Alpha as a person in an open-ended question.

Data analyses

SPSS 25 was used to test and answer the hypotheses and research questions. Univariate and multivariate outliers were examined using Box plots, Stem-and-leaf plots, and the Mahalanobis test. As each variable had less than 5% of missing data, the data can be considered missing at random (Schafer, 1999). Listwise deletion was used to deal with the missing data. Correlation, tolerance values, and variance inflation factor (VIF) values were used to test collinearity. Variables that were not normally distributed were log transformed to adjust the skewness and kurtosis. Three-way analyses of variance (ANOVAs) with vocal cues, kinetic cues, and gender as independent variables were conducted to test H1 and H2 and answer RQ1 and RQ2. RQ3 was answered by the three-way ANOVAs with vocal cues, kinetic cues, and robot use experiences as independent variables. The Process macro (Hayes, 2013) was used to conduct moderation analyses and examine RQ4. Model 1 and Johnson-Neyman technique were used in the moderation analyses.

Results

Manipulation checks were conducted using two-way ANOVAs. Those who were assigned to the human voice conditions ($M=5.96$, $SD=2.30$) reported the robot’s voice to be more humanlike, natural, and lifelike than those assigned to the synthetic voice conditions ($M=3.03$, $SD=2.09$), $F(1, 102)=46.37$, $p=.000$, partial $\eta^2=.31$. Those who were exposed to the robot’s gestural movements ($M=4.31$, $SD=2.03$) reported the robot’s movements to be more humanlike, natural, and lifelike than those exposed to non-gestural movements ($M=3.32$, $SD=2.03$), $F(1, 102)=6.47$, $p=.013$, partial $\eta^2=.06$. Therefore, the manipulation of the robot’s voices and movements was successful.

Table 1. The main effects of vocal and kinetic cues on users' social responses and anthropomorphism.

	Human voice	Synthetic voice	Main effects	Gestural movements	Non-gestural movements	Main effects
	<i>M (SD)</i>	<i>M (SD)</i>	<i>F</i>	<i>M (SD)</i>	<i>M (SD)</i>	<i>F</i>
Presence experience	6.94 (1.70)	6.71 (1.87)	0.54	6.88 (1.96)	6.77 (1.60)	0.03
Perceived attraction	5.87 (2.44)	5.43 (2.58)	1.48	6.40 (2.49)	4.88 (2.31)	10.28**
Perceived trustworthiness	8.42 (1.21)	7.50 (1.95)	10.00**	8.17 (1.54)	7.34 (1.81)	2.36
Intention of future use	7.82 (2.25)	7.64 (2.32)	0.47	8.23 (2.34)	7.21 (2.23)	5.57*
Physical anthropomorphism	5.96 (2.30)	3.03 (2.09)	46.37***	4.31 (2.03)	3.32 (2.03)	6.47*

* $p < .05$; ** $p < .01$; *** $p < .001$.

To test hypothesis 1, three-way ANOVAs suggested that the robot Alpha with a human voice ($M=8.42$, $SD=1.21$) evoked a significantly greater level of perceived trustworthiness than the one with a synthetic voice ($M=7.50$, $SD=1.95$), $F(1, 98)=10.00$, $p=.002$, partial $\eta^2=.09$. The robot's voice had the expected main effects on the perceived trustworthiness of the robot. Thus, H1(c) was supported. The robot with a human voice did not evoke greater levels of medium-as-social-actor presence, perceived attraction, and intention of future use than the one with a synthetic voice. H1(a), H1(b), and H1(d) were rejected.

To test hypothesis 2, three-way ANOVAs suggested that compared with non-gestural movements ($M=4.88$, $SD=2.31$), gestural movements ($M=6.40$, $SD=2.49$) led to significantly greater levels of perceived attraction of the robot, $F(1, 98)=10.28$, $p=.002$, partial $\eta^2=.10$. Gestural movements ($M=8.23$, $SD=2.34$) also led to significantly greater levels of intention of future use than non-gestural movements ($M=7.21$, $SD=2.23$), $F(1, 98)=5.57$, $p=.02$, partial $\eta^2=.05$. Thus, gestural movements had the expected main effects on perceived attraction of the robot and users' intention of future use. H2(b) and H2(d) were supported. However, compared with non-gestural movements, gestural movements did not lead to significantly greater levels of medium-as-social-actor presence or perceived trustworthiness of the robot. H2(a) and H2(c) were rejected. The results of H1, H2, and manipulation checks are shown in Table 1.

Research question 1 asked about the interaction effects between the social robot's vocal cues and kinetic cues. The three-way ANOVAs revealed no significant interaction between the vocal cues and the kinetic cues in predicting users' social responses. Research question 2 asked about the interaction effects between users' gender and the social cues in predicting users' social responses. The same three-way ANOVAs suggested that gender did not have main effects on the four types of social responses. However, gender interacted with the robot's kinetic cues in predicting users' intention of future use, $F(1, 98)=4.80$, $p=.031$, partial $\eta^2=.05$. The interaction suggested that compared with females, males reported greater levels of intention of future use when exposed to the robot's gestural movements. When exposed to non-gestural movements, females

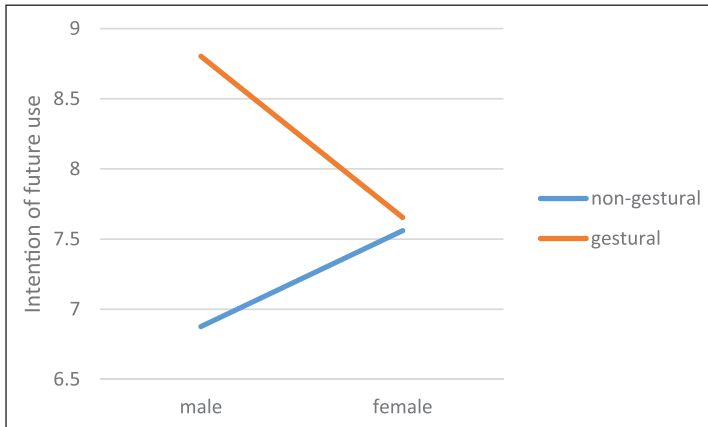


Figure 2. Interaction between gender and robot's movements on future use intention.

were more likely to report greater levels of intention of future use than males (Figure 2). Gender did not interact with vocal or kinetic cues in predicting other social responses.

Another set of three-way ANOVAs was conducted to examine how robot use experiences interacted with the social cues in predicting users' social responses (RQ3). Although there were no significant differences between those who had not interacted with any social robot in the previous year and those who had, robot use experiences interacted with the vocal cues in predicting medium-as-social-actor presence, $F(1, 98)=8.11, p=.005$, partial $\eta^2=.08$. Specifically, for those who had not interacted with any social robot in the previous year, robot Alpha's human voice evoked greater levels of medium-as-social-actor presence than its synthetic voice, while for those who had experiences using robots in the previous year, the human voice evoked lower levels of medium-as-social-actor presence than the synthetic voice (Figure 3).

Similarly, robot use experiences interacted with the vocal cues in predicting perceived attraction of the social robot, $F(1, 98)=5.12, p=.026$, partial $\eta^2=.05$, perceived trustworthiness of the robot, $F(1, 98)=9.19, p=.003$, partial $\eta^2=.09$, and users' intention of future use, $F(1, 98)=6.30, p=.014$, partial $\eta^2=.06$. That is, for those who had not interacted with any social robot in the previous year, robot Alpha's human voice evoked greater levels of perceived attraction, perceived trustworthiness, and intention of future use, while for those who had experiences using robots in the previous year, the human voice evoked lower levels of perceived attraction and intention of future use than the synthetic voice (see Figures 4 to 6). Robot use experiences did not interact with the kinetic cues in evoking users' social responses.

Moderation analyses were conducted to examine how users' attitudes toward robots interacted with the social cues in predicting users' social responses (RQ4). Results suggested that users' attitudes toward robots taking social roles did not interact with vocal cues in evoking users' social responses. However, there was a marginally significant interaction between the robot's kinetic cues and participants' attitudes toward robots in predicting medium-as-social-actor presence, $B=-.04, p=.0526$, lower limit of confidence interval

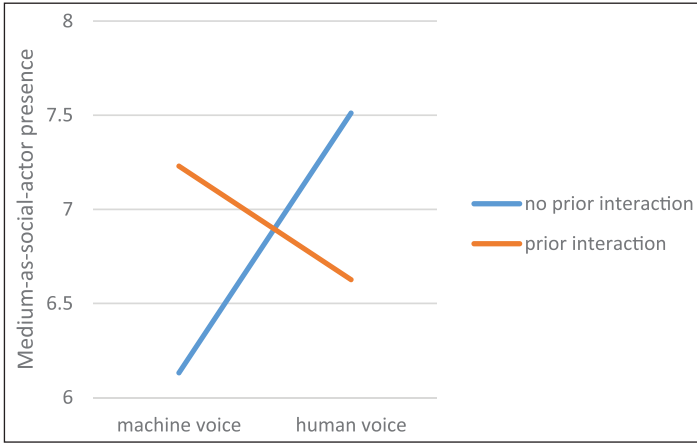


Figure 3. Interaction between robot use experiences and robot voices on medium-as-social-actor presence.

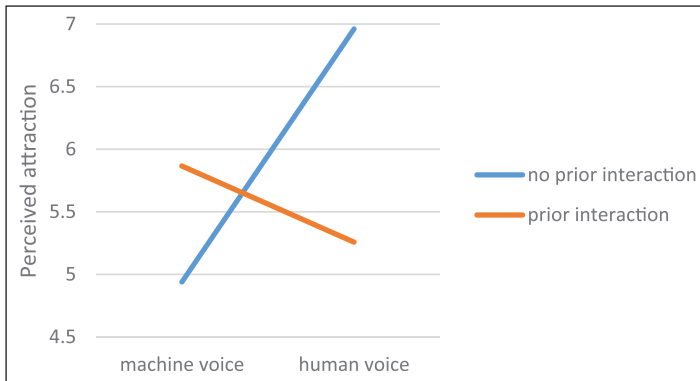


Figure 4. Interaction between robot use experiences and voices on perceived attraction.

(*LLCI*)=-0.09, upper limit of confidence interval (*ULCI*)=0.00. The results suggested that for those who felt comfortable with robots taking routinized social roles, the gestural movements of the social robot evoked greater levels of medium-as-social-actor presence than the non-gestural movements, while for those who were uncomfortable with robots taking routinized roles, gestural movements evoked lower levels of medium-as-social-actor presence than the non-gestural movements (see Figure 7).

Participants' attitude toward robots did not interact with the robot's kinetic cues in predicting perceived attraction, $B = .19, p > .05, LLCI = -0.33, ULCI = 0.71$, perceived trustworthiness of the robot, $B = .13, p > .05, LLCI = -0.23, ULCI = 0.49$, and intention of future use, $B = -.03, p > .05, LLCI = -0.09, ULCI = 0.04$. However, the Johnson-Neyman technique suggested that when the value of users' attitudes toward the robot

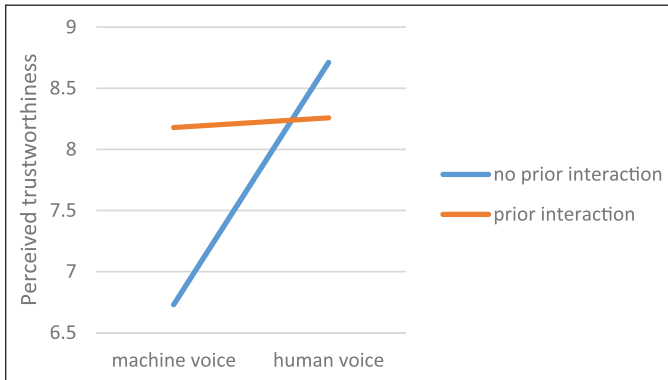


Figure 5. Interaction between robot use experiences and voices on perceived trustworthiness.

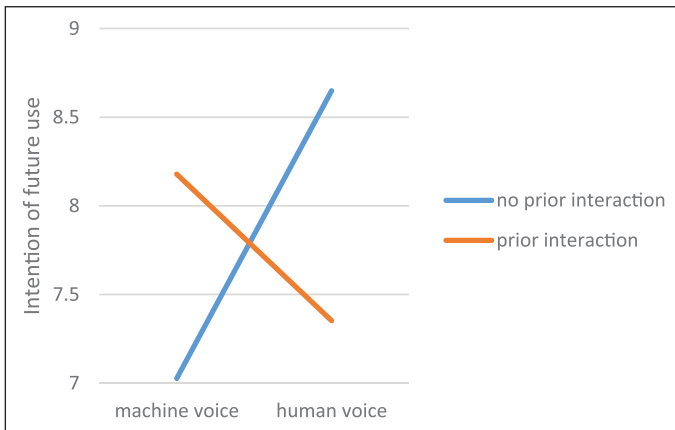


Figure 6. Interaction between robot use experiences and voices on intention of future use.

taking social roles was greater than 2.45, there was a significantly positive relationship between the robot’s gestures and perceived attraction of the robot. When the value of users’ attitudes toward robots taking social roles was greater than 2.85, there was also a positive relationship between the robot’s gestures and participants’ intention of future use of the robot. That is, generally for those who were comfortable with robots taking social roles in the society, the gestural movements of the robot led to more perceived attraction and intention of future use.

Discussion

This study focuses on users’ first encounter with the social robot Alpha and compares two pairs of social cues: human voice versus synthetic voice and gestural movements

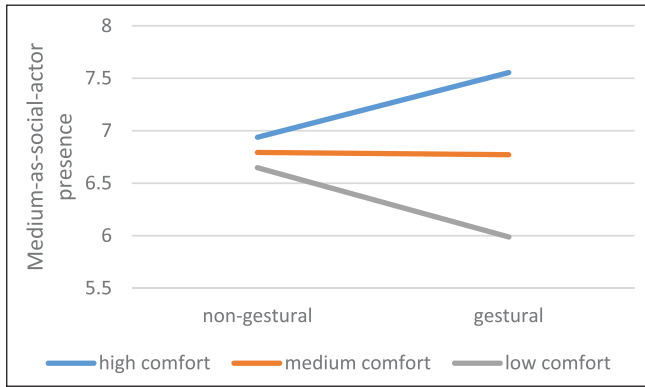


Figure 7. Interaction between the kinetic cues and participants' attitudes on presence. High comfort: 1 SD above the mean. Low comfort: 1 SD below the mean.

versus non-gestural movements. The results suggested that while the human voice of the robot increased its perceived trustworthiness, gestural movements heightened users' attachment to the robot and reinforced their future interaction motivations. The vocal cues and the kinetic cues of the robot did not have interaction effects, suggesting that the type of voice did not reinforce or reduce the effects of the robot's movements on users' social responses. In addition, inquiries into individual differences revealed that gender interacted with the robot's kinetic cues in users' intention of future use. Individuals' previous robot use experiences and their attitudes toward robots also leveraged their social responses to the social robot.

The power of the human voice over the synthetic voice in evoking users' trust supports the perspective that cues with more human features had stronger effects on users' social responses (Mayer et al., 2003; Tsimhoni et al., 2001). It is congruent with Fogg's (2002) finding that language with more anthropomorphic cues can increase the perceived credibility of computer technologies. However, vocal cues did not exert influence on users' medium-as-social-actor presence, perceived attraction, and intention of future use. The findings could be attributed to the perceived inconsistencies between the shape of the robot and its voice (Gong and Nass, 2007). If the participants perceived the robot's appearance to be machinelike, the consistency between their perception of the robot shape and the synthetic voice might have advanced users' social responses and counter-balanced the effects of the human voice on users' psychological reactions. This explanation has been corroborated by some participants' remarks to the open-ended question, where one participant noted, "I could see it with my own eyes and tell it was a robot and hear its robotic voice as well as robotic movements," even if that participant was assigned to the human voice and gestures condition.

Another reason for the non-significant differences in these types of social responses could be that participants' robot use experiences moderated the relationship between the vocal cues and social responses. For those who did not have any social robot use experiences in the previous year, the human voice was more likely to lead users to feel as if they were communicating with a social entity, feel intimate with the robot, and be

more willing to use a similar robot in the future, while for those who had experiences in interacting with social robots, the synthetic voice was more likely to evoke these social responses. The findings imply that more experienced robot users may feel more comfortable with a synthetic voice than a human voice. As scholars in prior research have only found either amplified effects (Johnson et al., 2004) or no effects of technology use experiences (Nass et al., 1995) on users' social responses to computer technologies, the current findings have added another layer to the literature where more technology use experiences could indeed undermine users' social responses to technologies. It may be because over time users tend to avoid uncanny valley effects (Mori et al., 2012), which refers to the idea that individuals could feel frightened when machines demonstrate rich social dimensions that blur the boundaries between humans and machines.

Gestural movements had the expected main effects on users' affinity for the social robot and intention of future use. The findings corroborate previous research on the positive effects of robots' gestures (Bevan and Fraser, 2015; Fiore et al., 2013). Nevertheless, gestural movements did not differ from non-gestural movements in affecting users' perception of the robot as a social actor. Based on the interaction effects, it might be because the effects of robots' gestural movements on users' medium-as-social-actor presence would be compromised when users were hesitant about what social roles robots should fill in the society, while those who were open to robots taking over routinized social roles preferred robots to present gestural movements.

Machine heuristics may also explain why gestural and non-gestural movements did not differ in predicting the credibility of the robot. It is possible that the robot's non-gestural movements may have triggered users' interpretation of the machine as unprejudiced and thus have advanced their trust to a similar level to the influence of gestural movements. Given the development of machine intelligence, this explanation may further raise new challenges to research on users' trust in robotic technologies as users' attitudes toward machines will become more intertwined and individualized. Future research could test the commonality and the differences between interpersonal trust and machine trust and conduct factor analyses to explore different dimensions of trust in HRI.

Although gender did not have main effect on users' social responses, results suggested that males were more likely than females to use a social robot in the future if the robot demonstrated gestures in its self-introduction. However, such gender differences should be further examined along with more complicated tasks and robots' actions (Kuchenbrandt et al., 2014). Considering that this study only manipulated the robot's movements for its self-introduction (e.g. waiving hands and bowing), it is overhasty to generalize these gender differences to other HRI contexts. As past research has revealed that males and females differ in their affective state based on the features of the HRI tasks they are assigned to (e.g. cooperative tasks vs competitive tasks; Mutlu et al., 2006), future research should factor in the effects of more specific robot actions and tasks to explicate gender differences.

Moreover, these gender differences in HRI could be shaped by individuals' cultural backgrounds. As people from different cultures may have different understandings of social roles such as babysitters, bosses, and judges, the extent to which males and females accept the robots with these social roles would be contingent upon the social norms. In past research, scholars have mostly focused on how culture could directly impact users'

attitudes toward social robots (Sabanovic, 2014). For instance, compared with US participants, Japanese participants reported more negative attitudes toward robots designed with emotional expressions (Bartneck et al., 2007). Nevertheless, little research has centered on the interaction between cultural influence and gender differences in users' acceptance of social robots. These studies could actually enrich our understanding of the role of gender in the human–robot relationship.

It should further be noted that robot Alpha was designed with a male voice to match its masculine appearance. Therefore, the male voice itself could be a confounding factor in the experiment. Hence, scholars could test the effects of social robots with more gender representations (Carpenter et al., 2009). Considering that current robotic technologies may reflect and even intensify the existent prejudices in the society, researchers should use the design opportunities to understand and reduce the stereotyped characteristics of these technologies (AI Now Institute, 2018).

When participants were asked whether and why they perceived the robot as a person, most of them explicitly denied perceiving it as a social actor. Some participants explained that it was because the robot exhibited insufficient human characteristics and made disturbing mechanical noises. One participant wrote, "One thing that stuck out through its introduction was the noise that it made whenever it moved, which at times was louder than its voice, which made it seem more robotic and less person-like." A few participants emphasized that they were clear about the nature of the robot not being a person from the beginning of the experiment and were unwilling to suspend their disbelief. One participant assigned to the human voice and gestural movements condition wrote, "Although I was impressed and intrigued by Alpha's humanlike behavior, I could not get past the feeling that he is just a machine created by humans."

While the majority of the participants denied perceiving the robot as a person, the mean values of participants' responses to the measures of medium-as-social-actor presence, perceived trustworthiness, and intention of future use in each condition of the experiment were all over six on the 10-point scales, suggesting that even though participants did not believe that humanoid robots warrant human treatment, they applied social scripts in communication with the robot Alpha "without extensive thought or deliberation" (Moon, 2000: 325). The finding can support the mindlessness explanation (Nass and Moon, 2000). It is also compatible with Kim and Sundar's (2012) finding that users intentionally denied treating a website with a humanlike agent in social manners, but mindlessly reported the website to be more friendly, personal, sociable, and likable.

Notwithstanding the mindlessness explanation, it is still premature to conclude that users respond to humanoid social robots without controlled processing (Langer, 1992). Indeed, a few participants acknowledged their perception of the robot as a person. One participant described, "When Alpha first started talking and I heard a sound of silence, I began to talk back as if we were engaged in a conversation." Another participant also felt communicating with a social actor, "This is because the robot Alpha spoke directly to me, introduced himself to me, and was able to move right in front of my eyes, all qualities of a human being." However, considering that these positive responses were self-reported and could only be regarded as indirect evidence of users' psychological processes, scholars should continue to use other methods to identify through which

cognitive route users react to these technologies (Lee, 2010). For instance, physiological measures may help researchers understand users' brain activities during their interactions with robotic technologies. Researchers could also manipulate the wordings in their questionnaire items to see how users report their reactions differently.

This study has both theoretical and practical implications. First, this study responds to Nass and Moon's (2000) call for more CASA research on the comparisons of different social dimensions of computer technologies. The study demonstrates that each single social cue has its unique power in evoking users' social perception and social attitudes. The findings can contribute to the HRI literature by illustrating the connections between the richness of the cues and users' social responses to technologies. Future research could compare more pairs of cues and develop a hierarchy of social cues that explain their distinctive impacts on users' mental and behavioral reactions. These social cues could be non-verbal physical ones such as human shape and human size as well as abstract human characteristics such as interactivity cues, identity cues, and emotional cues.

Second, this study serves as an additional example of how the CASA paradigm can be applied to not only computers but also humanoid social robots. Compared with some of the recent studies that have employed the paradigm to investigate the perceived personalities of zoomorphic robots, typos produced by online chatbots, and haptic cues of social robots (Lee et al., 2006; Li et al., 2017; Spence et al., 2014; Westerman et al., 2018), this study illustrates the potential of applying the CASA paradigm to understand users' first impression of a humanoid social robot that delivers both vocal and kinetic messages in a socioemotional context.

Third, this study suggests that when expanding the CASA paradigm, researchers should take both the quality of social cues and the individual differences into consideration. For example, participants' open-ended remarks implied that personal expectations for robot performances could be influential in HRI. One participant explicitly denied perceiving the robot as a social actor as the robot did not "speak like Siri or Alexa." Popular media portrayals may also have forged people's imagination of social robots and swayed users' attitudes toward them. This is aligned with Paepcke and Takayama's (2010) finding that participants perceived a robot as less competent when they had high expectations of its interactive touch-sensing capabilities.

Beyond personal expectations, how individuals self-identified their age could affect their technology use experiences. Edwards et al. (2019) found that users in a high age identification group evaluated an older AI voice to be more credible and more socially present than those in a low age identification group. Guzman (2019) explored users' interactions with different mobile virtual assistants and argued that individuals may differ in orienting the voices to be in the smartphone or of the smartphone. These studies have laid out more possibilities to include individual differences in HRI.

This study also has practical implications. As human voices have more straightforward effects on perceived trustworthiness of the robot than gestural movements, designers could embed human voices to social robots when they are designed to deliver persuasive information for educative purposes, health care recommendations, or business negotiations. If the goal of human-robot interaction is to increase their socioemotional connection, then designing gestural movements into the robot should be emphasized, as gestures would be much more powerful in evoking users' affinity for and

acceptance of the robot. Those who design the kinetic cues and vocal cues into robots should also customize their products based on personal experiences and preferences. For example, compared with a human voice, a synthetic voice could be prioritized for more experienced robot users.

While customizing these social cues may enhance user experiences, designers should understand the differences between people's social use and functional use of technologies. Although some industrial robots and automated machines (e.g. washing machines and microwaves) are designed with human-centered interfaces, the major functions of these technologies are not for humanlike social interactions (Zhao, 2006). The differences between social robots and these automated machines may guide researchers to apply different design principles. For example, when designing machines that have utilitarian purposes (e.g. robot vacuum cleaners and self-piloting cars), researchers should base their design on the goal of increasing humans' efficiency of communication with machines. Designers could make the verbal or text-based instructions from the user interface more straightforward, explicit, and instrumental. When users operate or respond to the human cues of these machine interfaces, a certain level of social responses will be evoked in the process, but the degree of these social responses does not need to be as strong and broad as that in human interactions. Although nowadays the boundary between social robots and industrial robots have become increasingly vague, in the process of optimizing technology design, researchers should fathom to what degree users desire or fear social cues in both short-term and long-term human–robot relationships.

In addition to these theoretical and practical implications, researchers should be fully aware of the potential ethical risks of applying the findings to technology innovation. It would be perilous to increase the perceived trustworthiness of a robot if it is used to deliver fake news. As potential solutions, scholars or experts should formulate and adopt comprehensive ethical codes to prevent potential harm on technology users (Lombard, 2010).

Conclusion

Today, technology advancements have made social robots more accessible than ever before. However, what may be the most effective and efficient way to interact with these robots remains to be explored. To better understand the human–robot relationship, this study examines users' social responses to a social robot through the lens of the quality of social cues. Centering on the interaction effects between technological features and individual differences, this study has provided a prospect for expanding and updating the CASA paradigm to explain HRI. The study has not only revealed some social and design implications for human–machine communication but also pointed the way to more nuanced and comprehensive discoveries of users' psychological processing of social robots in the future.

Meanwhile, this study is not without limitations. First, when asked about robot use experiences, participants were provided with Zhao's (2006) definition of humanoid social robots, which were referred to as "human-made autonomous entities that interact with humans in a humanlike way" (p. 405). However, this definition of humanoid social robots includes not only physically embodied human-looking entities but also technologies such as virtual assistants or chatbots. This broad definition may have led participants

to report use experiences of a wide range of robotic technologies. Therefore, in future research, scholars should be more accurate about which type of robots they expect participants to think of, as robots such as Aibo, NAO, Alexa, and Jibo vary in their shapes, motion, presence, and affordances.

Second, this study focuses on participants' first encounter with the robot Alpha. Although some participants said "hi" and "how are you" back to the robot when the robot greeted them during self-introduction, they were not instructed to formally interact with the robot Alpha. More research could be conducted to test the natural conversations between humans and social robots.

Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

ORCID iD

Kun Xu  <https://orcid.org/0000-0001-9044-821X>

References

- AI Now Institute (2018) Bias and inclusion. Available at: <https://ainowinstitute.org/research.html>
- Bartneck C and Forlizzi J (2004) A design-centered framework for social human-robot interaction. In: *Proceedings of the 13th IEEE international workshop on robot and human interactive communication*, Kurashiki, Japan, 22 September, pp. 591–594. New York: IEEE.
- Bartneck C, Reichenbach J and Carpenter J (2008) The carrot and the stick: the role of praise and punishment in human-robot interaction. *Interaction Studies* 9: 179–203.
- Bartneck C, Suzuki T, Kanda T, et al. (2007) The influence of people's culture and prior experiences with Aibo on their attitude towards robots. *AI & Society* 21: 217–230.
- Bevan C and Fraser DS (2015). Shaking hands and cooperation in tele-present human-robot negotiation. In: *Proceedings of the 10th annual ACM/IEEE international conference on human-robot interaction*, Portland, OR, 2–5 March, pp. 247–254. New York: ACM.
- Biocca F, Harms C and Burgoon JK (2003) Toward a more robust theory and measure of social presence: review and suggested criteria. *Presence: Teleoperators and Virtual Environments* 12(5): 456–480.
- Bracken CC and Lombard M (2004) Social presence and children: praise, intrinsic motivation, and learning with computers. *Journal of Communication* 54(1): 22–37.
- Breazeal C (2003) Emotion and sociable humanoid robots. *International Journal of Human-Computer Studies* 59(1): 119–155.
- Cabibihan JJ, So WC and Pramanik S (2012) Human-recognizable robotic gestures. *IEEE Transactions on Autonomous Mental Development* 4(4): 305–314.
- Carpenter J, Davis JM, Erwin-Stewart N, et al. (2009) Gender representation and humanoid robots designed for domestic use. *International Journal of Social Robotics* 1: 261–265.
- Carr CT, Vitak J and McLaughlin C (2011) Strength of social cues in online impression formation: expanding SIDE research. *Communication Research* 40: 261–281.
- Copleston SN and Bugmann G (2008) Personal robot user expectations. *Advances in Communications, Computing, Networks, and Security* 5: 230–238.
- Derks D, Bos AER and Von Grombkow J (2007) Emoticons and social interaction on the Internet: the importance of social context. *Computers in Human Behavior* 23: 842–849.

- Duffy BR (2003) Anthropomorphism and the social robot. *Robotics and Autonomous Systems* 42: 177–190.
- Duquette A, Michaud F and Mercier H (2008) Exploring the use of a mobile robot as an imitation agent with children with low-functioning autism. *Autonomous Robots* 24: 147–157.
- Edwards C, Edwards A, Spence PR, et al. (2014) Is that a bot running the social media feed? Testing the differences in perceptions of communication quality for a human agent and a bot agent on Twitter. *Computers in Human Behavior* 33: 372–376.
- Edwards C, Edwards A, Stoll B, et al. (2019) Evaluations of an artificial intelligence instructor's voice: social identity theory in human-robot interactions. *Computers in Human Behavior* 90: 357–362.
- Engelbart DC (1962) *Augmenting human intellect: a conceptual framework*. SRI Project 3578 for Air Force Office of Scientific Research. Stanford Research Institute. Available at: <http://dougengelbart.org/content/view/138>
- Eyssel F and Kuchenbrandt D (2012) Social categorization of social robots: anthropomorphism as a function of robot group membership. *British Journal of Social Psychology* 51: 724–731.
- Fiore SM, Wiltshire TJ, Lobato EJC, et al. (2013) Toward understanding social cues and signals in human-robot interaction: effects of robot gaze and proxemics behavior. *Frontiers in Psychology* 4: 1–15.
- Fogg BJ (2002) Computers as persuasive social actors. In: Fogg BJ (ed.) *Persuasive Computers Using Technology to Change What We Think and Do*. San Francisco, CA: Morgan Kaufman Publishers, pp. 89–120.
- Gannon M (2013) Japan's Kirobo talking robot chats with astronaut in Space for 1st time. *Space*, December. Available at: <http://www.space.com/24044-kirobo-talking-robot-astronaut-chat-video.html>
- Gong L (2008) How social is social responses to computers: the function of the degree of anthropomorphism in computer representations. *Computers in Human Behavior* 24: 1494–1509.
- Gong L and Lai J (2003) To mix or not to mix synthetic speech and human speech? Contrasting impact on judge-rated task performance versus self-rated performance and attitudinal responses. *International Journal of Speech Technology* 6: 123–131.
- Gong L and Nass C (2007) When a talking face computer agent is half-human and half-humanoid: human identity and consistency preference. *Human Communication Research* 33: 163–193.
- Gunawardena CN and Zittle FJ (1997) Social presence as a predictor of satisfaction within a computer-mediated conferencing environment. *American Journal of Distance Education* 11: 8–26.
- Guzman A (2019) Voices in and of the machine: source orientation toward mobile virtual assistants. *Computers in Human Behavior* 90: 343–350.
- Hayes AF (2013) *Introduction to Mediation, Moderation, and Conditional Analysis: A Regression-based Approach*. New York: Guilford Press.
- Heider F and Simmel M (1944) An experimental study of apparent behavior. *American Journal of Psychology* 57: 243–259.
- Herring SC (2004) Slouching toward the ordinary: current trends in computer-mediated communication. *New Media & Society* 6(1): 26–36.
- Hinds PJ, Roberts TL and Jones H (2004) Whose job is it anyway? A study of human-robot interaction in a collaborative task. *Human-computer Interaction* 19: 151–181.
- Hoffman G and Ju W (2012) Designing robots with movement in mind. *Journal of Human Robot Interaction* 1(1): 78–95.
- Johansson G (1973) Visual perception of biological motion and a model for its analysis. *Perception & Psychophysics* 14: 201–211.
- Johnson D, Gardner J and Wiles J (2004) Experience as a moderator of the media equation: the impact of flattery and praise. *International Journal of Human-Computer Studies* 61: 237–258.

- Ju W and Sirkin D (2010) Animate objects: how physical motion encourages public interaction. In: *Proceedings of the international conference on persuasive technology*, Copenhagen, 7–10 June, pp. 40–51. Berlin: Springer.
- Ju W and Takayama L (2009) Approachability: how people interpret automatic door movement as gesture. *International Journal of Design* 3: 1–10.
- Jung M, Martelero N, Hoster H, et al. (2014) Participatory materials: having a reflective conversation with an artifact in the making. In: *Proceedings of the 2014 conference on designing interactive systems*, Vancouver, BC, Canada, 21–25 June, pp. 25–34. New York: ACM.
- Jung Y and Lee KM (2004) Effects of physical embodiment on social presence of social robots. In: *Proceedings of PRESENCE*, pp. 80–87. Available at: <https://pdfs.semanticscholar.org/6efb/d084bb99e018a82ed6823fa8ec1c9b7cfc38.pdf>
- Kim A, Han J, Jung Y, et al. (2013) The effects of familiarity and robot gesture on user acceptance of information. In: *Proceedings of the 2013 8th ACM/IEEE international conference on human-robot interaction (HRI)*, Tokyo, Japan, 3–6 March, pp. 159–160. New York: IEEE.
- Kim Y and Sundar SS (2012) Anthropomorphism of computers: is it mindful or mindless? *Computers in Human Behavior* 28: 241–250.
- Krauss RM, Chen Y and Chawla P (1996) Nonverbal behavior and nonverbal communication: what do conversational hand gestures tell us? *Advances in Experimental Social Psychology* 28: 389–450.
- Kuchenbrandt D, Haring M, Eichberg J, et al. (2014) Keep an eye on the task! How gender typicality of tasks influence human-robot interactions. *International Journal of Social Robotics* 6: 417–427.
- Langer EJ (1992) Matters of mind: mindfulness/mindlessness in perspective. *Consciousness and Cognition* 1: 289–305.
- Lee EJ (2008) Flattery may get computers somewhere, sometimes: the moderating role of output modality, computer gender, and user gender. *International Journal of Human-computer Studies* 66: 789–800.
- Lee EJ (2010) What triggers social responses to flattering computers? Experimental tests of anthropomorphism and mindlessness explanations. *Communication Research* 37: 191–214.
- Lee KM (2004) Presence explicated. *Communication Theory* 14(1): 27–50.
- Lee KM, Park N and Song H (2005) Can a robot be perceived as a developing creature? Effects of a robot's long-term cognitive developments on its social presence and people's social responses toward it. *Human Communication Research* 31: 538–563.
- Lee KM, Peng W, Jin SA, et al. (2006) Can robots manifest personality? An empirical test of personality recognition, social responses, and social presence in human-robot interaction. *Journal of Communication* 56: 754–772.
- Li J (2015) The benefit of being physically present: a survey of experimental works comparing copresent robots, telepresent robots and virtual agents. *International Journal of Human-computer Studies* 77: 23–37.
- Li J, Ju W and Reeves B (2017) Touching a mechanical body: tactile contact with intimate parts of a humanoid robot is physiologically arousing. *Journal of Human-Robot Interaction* 6: 118–130.
- Li J, Kizilcec R, Bailenson J, et al. (2015) Social robots and virtual agents as lecturers for video instruction. *Computers in Human Behavior* 55: 1222–1230.
- Licklider JC and Taylor RW (1968) The computer as a communication device. *Science and Technology* 76(2): 1–3.
- Lombard M (2010) The promise and peril of telepresence. In: Bracken CC and Skalski PD (eds) *Immersive in Media: Telepresence in Everyday Life*. New York: Routledge, pp. 197–218.

- Lombard M and Ditton T (1997) At the heart of it all: the concept of presence. *Journal of Computer-Mediated Communication* 30: 184–198.
- McCroskey J and McCain T (1974) The measurement of interpersonal attraction. *Speech Monographs* 41: 261–266.
- Machotka P (1965) Body movement as communication. *Dialogues: Behavioral Science Research* 2: 33–66.
- Mayer RE, Sobko K and Mautone P (2003) Social cues in multimedia learning: role of speaker's voice. *Journal of Educational Psychology* 95: 419–425.
- Mehrabian A (1969) Encoding of attitude by a seated communicator via posture and position cues. *Journal of Consulting and Clinical Psychology* 33: 330–336.
- Moon Y (2000) Intimate exchanges: using computers to elicit self-disclosure from consumers. *Journal of Consumer Research* 26: 323–339.
- Mori M, MacDorman KF and Kageki N (2012) The uncanny valley. *IEEE Robotics & Automation Magazine* 19: 98–100.
- Mutlu B, Osman S, Forlizzi J, et al. (2006) Perceptions of ASIMO: an exploration on cooperation and competition with humans and humanoid robots. In: *Proceedings of the 1st ACM SIGCHI/SIGART conference on human-robot interaction*, Salt Lake City, UT, 2–3 March, pp. 351–353. New York: ACM.
- Nass C (2004) Etiquette quality: exhibitions and expectations of computer politeness. *Communications of the ACM* 47: 35–37.
- Nass C and Brave S (2005) *Wired for Speech: How Voice Activates and Advances the Human-Computer Relationship*. Cambridge, MA: The MIT Press.
- Nass C and Lee KM (2001) Does computer-synthesized speech manifest personality? Experimental tests of recognition, similarity attraction, and consistency attraction. *Journal of Experimental Psychology: Applied* 7: 171–181.
- Nass C and Moon Y (2000) Machines and mindlessness: social responses to computers. *Journal of Social Issues* 56: 81–103.
- Nass C and Steuer J (1993) Voices, boxes, and sources of messages: computers and social actors. *Human Communication Research* 19: 504–527.
- Nass C, Fogg BJ and Moon Y (1996) Can computers be teammates? *International Journal of Human Computer Studies* 45: 669–678.
- Nass C, Lombard M, Henriksen L, et al. (1995) Anthropocentrism and computers. *Behavior & Information Technology* 14: 229–238.
- Nass C, Moon Y and Green N (1997) Are machines gender neutral? Gender-stereotypic responses to computers with voices. *Journal of Applied Social Psychology* 27: 864–876.
- Nass C, Steuer J and Tauber ER (1994) Computers are social actors. *Human Factors in Computing Systems* 94: 72–78.
- National Institute of Health (NIH) (2017) What is voice? What is speech? *What is language?* March. Available at: <https://www.nidcd.nih.gov/health/what-is-voice-speech-language>
- Nomura T, Kanda T, Suzuki T, et al. (2009) Age differences and images of robots: social survey in Japan. *Interaction Studies* 10: 374–391.
- Norman D (1988) *The Psychology of Everyday Things*. New York: Basic Books.
- Oh CS, Bailenson JN and Welch GF (2018) A systematic review of social presence: definition, antecedents, and implications. *Frontiers in Robotics and AI* 5: 1–34.
- Paepcke S and Takayama L (2010) Judging a bot by its cover: an experiment on expectation setting for personal robots. In: *Proceedings of the 5th ACM/IEEE international conference on Human-robot interaction*, Osaka, Japan, 2–5 March, pp. 45–52. New York: IEEE.
- Pfeifer R and Scheier C (1999) *Understanding Intelligence*. Cambridge, MA: The MIT Press.

- Picard RW (1995) Affective computing. *MIT Media Laboratory Perceptual Computing Section Technical Report 321*: 1–26.
- Postmes T, Spears R and Lea M (1998) Breaching or building social boundaries? Side-effects of computer-mediated communication. *Communication Research* 25: 689–715.
- Powers A and Kiesler S (2006) The advisor robot: tracing people’s mental model from a robot’s physical attributes. In: *Proceedings of the 1st ACM SIGCHI/SIGART conference on human-robot interaction*, March, pp. 218–225. New York: ACM.
- Reeves B and Nass C (1996) *The Media Equation: How People Treat Computers, Television, and New Media Like Real People and Places*. Stanford, CA: CSLI Publications.
- Sabanovic S (2014) Inventing Japan’s robotics culture: the repeated assembly of science, technology, and culture in social robotics. *Social Studies of Science* 44: 342–367.
- Sah YJ and Peng W (2015) Effects of visual and linguistic anthropomorphic cues on social perception, self-awareness, and information disclosure in a health website. *Computers in Human Behavior* 45: 392–401.
- Salem M, Eyssel F, Rohlffing K, et al. (2013) To err is human(-like): effects of robot gesture on perceived anthropomorphism and likability. *International Journal of Social Robot* 5: 313–323.
- Salem M, Kopp S, Wachsmuth I, et al. (2010) Towards an integrated model of speech and gesture production for multi-modal robot behavior. In: *Proceedings of the 19th international symposium in robot and human interactive communication*, Viareggio, 13–15 September.
- Salem M, Rohlffing K, Kopp S, et al. (2011) A friendly gesture: investigating the effect of multimodal robot behavior in human-robot interaction. In: *Proceedings of the 2011 RO-MAN*, Atlanta, GA, 31 July–3 August.
- Schafer JL (1999) Multiple imputation: a primer. *Statistical Methods in Medical Research* 8(1): 3–15.
- Searle JR (1980) Minds, brains, and programs. *Behavioral and Brain Sciences* 3: 417–424.
- Shin DH and Choo H (2011) Modeling the acceptance of socially interactive robotics: social presence in human–robot interaction. *Interaction Studies* 12(3): 430–460.
- Sirkin D and Ju W (2012). Consistency in physical and on-screen action improves perceptions of telepresence robots. In: *Proceedings of the 7th annual ACM/IEEE international conference on human-robot interaction*, Boston, MA, 5–8 March, pp. 57–64. New York: ACM.
- Spence PR, Westerman D, Edwards C, et al. (2014) Welcoming our robot overloads: initial expectations about interaction with a robot. *Communication Research Reports* 31: 272–280.
- Stoll B, Edwards C and Edwards A (2016) “Why aren’t you a sassy little thing?": the effects of robot-enacted guilt trips on credibility and consensus in a negotiation. *Communication Studies* 67(5): 530–547.
- Suchman LA (2007) *Human-Machine Reconfigurations: Plans and Situated Actions*. Cambridge: Cambridge University Press.
- Sundar S (2008) The MAIN model: a heuristic approach to understanding technology effects on credibility. In: Metzger MJ and Flanagin AJ (eds) *Digital Media, Youth, and Credibility*. Cambridge, MA: The MIT Press, pp. 73–100.
- Sundar S and Nass C (2001) Conceptualizing sources in online news. *Journal of Communication* 51(1): 52–72.
- Sundar SS, Xu Q and Oeldorf-Hirsch A (2009) Authority vs. peer: how interface cues influence users. In *Proceedings of the CHI '09 extended abstracts on human factors in computing systems*, Boston, MA, 4–9 April, pp. 4231–4236. New York: ACM.
- Tsimhoni O, Green P and Lai J (2001) Listening to natural and synthesized speech while driving: effects on user performance. *International Journal of Speech Technology* 4(2): 155–169.

- Tung FW and Deng YS (2007) Increasing social presence of social actors in e-learning environments: effects of dynamic and static emoticons on children. *Displays* 28: 174–180.
- Turing AM (1950) Computing machinery and intelligence. *Mind* 59: 433–460.
- UBTECH (2017) Alpha 1S: the fun family robot, March. Available at: <http://www.ubtrobot.com/index.html>
- Wang LC, Baker J, Wagner JA, et al. (2007) Can a retail web site be social? *Journal of Marketing* 71(3): 143–157.
- Weiser M (1991) The computer for the 21st century. *Scientific American* 265: 94–110.
- Westerman D, Cross AC and Lindmark PG (2018) I believe in a thing called bot: perceptions of the humanness of chatbots. *Communication Studies*. Epub ahead of print 20 December. DOI: 10.1080/10510974.2018.1557233.
- Wiener N (1950) *Human Use of Human Beings*. Cambridge: Da Capo Press.
- Xu K and Lombard M (2016) Media are social actors: expanding the CASA paradigm in the 21st Century. *Paper presented at the Annual Conference of the International Communication Association*. Fukuoka, Japan.
- Zhao S (2003) Toward a taxonomy of copresence. *Presence* 12: 445–455.
- Zhao S (2006) Humanoid social robots as a medium of communication. *New Media & Society* 8: 401–419.

Author biography

Kun Xu (PhD, Temple University, 2018) is an assistant professor in the Department of Telecommunication at University of Florida. His research investigates how people interact with and respond to social robots, how people use emerging technologies to experience virtual and augmented spaces, and how the evolution of media environments affects people's cognition, emotion, and behavior.