

Contents lists available at ScienceDirect

# International Journal of Human - Computer Studies

journal homepage: www.elsevier.com/locate/ijhcs



# Check for updates

# Social interactions and relationships with an intelligent virtual agent

Sun Kyong Lee\*, Pavitra Kavya, Sarah C. Lasser

Department of Communication, The University of Oklahoma, 610 Elm Ave., Norman, OK 73019, United States

#### ARTICLE INFO

Keywords:
Intelligent virtual agent
Human-computer trust
Task type
Gender
Copresence
faith
Personal attachment
Siri as a coworker
Siri as a supervisor
and Siri as a friend

#### ABSTRACT

The current study explores how humans interact with an intelligent virtual agent (IVA), Siri, on their mobile phones and how such experiences may impact their trust, social presence, and comfort level toward the IVA as a coworker, supervisor, and friend. A two (male- vs. female-voice of Siri) by two (functional vs. social task) by two (matched vs. ummatched with participants' gender) between-subjects experiment was performed with 163 participants. Higher levels of trust in cognitive dimensions were associated with functional tasks assisted by Siri, and one interaction between participants' gender and Siri's gendered voice was observed in an affective dimension of trust, faith. Copresence dimension of social presence was associated with technical competence, understandability, faith, and personal attachment dimension of trust whereas psychological involvement was related with reliability and technical competence. Copresence, alongside faith and personal attachment, both affective dimensions of trust, seemed to facilitate comfortable feelings toward Siri imagined as various relational partners.

#### 1. Introduction

Intelligent virtual agents (IVAs) powered by artificial intelligence (AI) are prevalent in our daily lives. Amazon's Alexa, Apple's Siri, and Microsoft's Cortana are all IVAs that search information in real-time and verbally communicate the results to humans (Sin & Munteanu, 2020). The current study explores how humans interact with an IVA, Siri, on their mobile phones and how those experiences may influence their attitudes toward Siri as a coworker, supervisor, and friend. As people's interpretations of a communication technology influence their opinions of and behaviors toward it (Sundar, 2008), examining how they interpret the functions and utility of IVAs is important to understanding their current and near future evaluations of the agents.

Experiences of interacting with communicative IVAs are already shaping peoples' attitudes toward more embodied AI-agents, like social robots (Guzman, 2017). Social robots, with whom humans can interact and form relationships similar to those they foster with other humans, are not yet so common; but various forms of other robots are being actively developed, tested, and utilized in the military, medical surgery, and commercial businesses. The future of society may involve routine assistance from robots and other AI-systems in almost all areas of life, and with humans living and working side-by-side with them. Thus, whether and how humans trust robots and AI-systems has great implications for their coexistence. Though envisioned by many sci-fi films for

decades, the potential for this type of robot-assisted society to become a near-future reality is greater than ever.

Based on an in-lab experiment, we first tested how Siri's genderedvoice (male vs. female), different types of tasks (functional vs. social), and participants' gender match with Siri's gendered-voice influenced their trust of the IVA and social presence felt during an interaction episode. Motivated by the fact many commercially available IVAs' default voices are female (e.g., Alexa, Siri, and Cortana), but can be changed to a male's voice, and by previous research showing gender stereotypes (Damen & Toh, 2019; Guzman, 2017; Nass et al., 1997) and social identification (gender match; Lee et al., 2000; Payne et al., 2013; Ruijten et al., 2015) effects in human-computer (robot) interactions, this study used Siri's gendered voice and its match with the participant's as one of the major components of experimental stimuli. In addition, motivated by the increasing potential of humans' forming meaningful relationships with AI-systems including robots, the study included functional as well as social tasks in which humans can engage an IVA. Performing functional tasks assisted by an automated machine is certainly more common than social or socializing tasks; however, as many machines are developed to have anthropomorphic features (e.g., voices, appearances) and humans may perceive them as not only tools, but also social beings, it is worthwhile to examine whether distinct task types bring differential impact on humans' trust of an IVA.

Second, while controlling for the impact of gendered voice, task type,

E-mail addresses: sunklee@ou.edu (S.K. Lee), pkavya@cpp.edu (P. Kavya), sarah.lasser@ou.edu (S.C. Lasser).

<sup>\*</sup> Corresponding author.

gender match, and participant's proficiency and length of Siri usage, we examined how social presence and five-dimensions of human-computer trust (i.e., reliability, technical competence, understandability, faith, and personal attachment) were associated with their attitudes towards Siri as a coworker, supervisor, or friend. Social presence has been found to be an important mediating factor in human-computer (robot) interactions (Fortin & Dholakia, 2005; Gefen & Straub, 2004; Hassanein & Head, 2006; Karahanna & Straub, 1999). Whether human users experience their interactions with IVAs as being similar to those with other humans would have implications for their trust of the AI-system. Forming trust might be easier for humans if they perceive high levels of social presence, in other words feeling the interaction with an IVA as lively and attentive as an interaction with another human being.

Findings of this study contribute to theory building in the burgeoning area of human-machine communication (HMC). While differentiating the field from more established human-computer interaction (HCI) research, Guzman and Lewis (2019) called for HMC research in three areas: functional, relational, and metaphysical. The current study addresses the functional aspects of HMC by examining the relationships between the nature of task and humans' trust and social presence toward an IVA. Furthermore, this study investigates how humans react to the potential scenario of working with the IVA in various relational arrangements (e.g., coworker, supervisor, or friend). Results shed light on the current state of HMC and HCI research and expand knowledge about task and gender effects of IVA. After a review of the relevant literature and theories, hypotheses and research questions of the study will be suggested.

#### 2. Literature Review

#### 2.1. Human-Computer Trust and Task

As a wide range of previously human-performed tasks become automated and decision-making evolves via assistance from AI-agents, trust in human-computer interactions becomes increasingly important (Lee & Nass, 2010; J. D. Lee & See, 2004; Madsen & Gregor, 2000; Yagoda & Gillan, 2012). For the scope of this study, trust is defined as "the attitude that an agent will help achieve an individual's goals in a situation characterized by uncertainty and vulnerability" (Hancock et al., 2011; Lee & See, p. 54). An agent here can be a computer, an IVA, or an AI-robot. Considering the possibility that humans may regularly work with AI-agents such as robots as coworkers in the near future, and trust in IVAs can influence other attitudes (e.g., comfort, dependence) toward robots, additional research is necessary to unpack the various dimensions of trust associated with humans' interactions with AI-agents.

Trust between humans and AI-agents shapes individuals' ability to acknowledge information and follow suggestions from robots in critical situations like combat missions (Groom & Nass, 2007). The results from a meta-analysis suggest that among multiple factors that impact humans' relationships with robots, a robot's performance is the primary factor moderating trust: a higher level of trust is associated with a more reliable performance by a robot (Hancock et al., 2011). Working with AI-agents such as robots can augment human capabilities; for example, humans can have robots take care of mundane, and sometimes dangerous, tasks while engaging in more creative, critical, and complex tasks. With the advancement of AI, robotic technology, and autonomous unmanned vehicles, human-computer trust becomes a paramount issue (Lee & Nass, 2010; Yagoda & Gillan, 2012).

Much research has applied the computers-are-social-actors (CASA) perspective to examine which human and social factors are associated with users' perception of computer agents' trustworthiness (Lee & Nass, 2010). Reeves and Nass (1996) proposed the CASA paradigm as an extension of their media equation theory that "a computer is fundamentally social and humans apply wide sets of social characteristics to a computer when the computer manifests human-like characteristics such as language, social roles, gender, ethnicities, and personality" (Lee et al.,

2005, p. 540). In their experiments, people were polite to computer-generated voices and more favorable to the voices that flattered them. Existing gender stereotypes applied when interacting with female- vs. male-voiced computers (Lee et al., 2000; Nass et al., 1997). Reeves and Nass, relying on an evolutionary perspective, explained humans' brain have not evolved to deal with interactions with new communicative objects like computers, so our application of social rules to interaction with machines is automatic and natural.

Research applying CASA to the study of trust found humans tended to trust a computerized voice that matched their own personality (extroversion vs. introversion; Nass & Lee, 2001). Although artificial, humans trusted computers with emphatic facial images more in competitive game situations (Brave et al., 2005), and learning companion agents who provided supportive messages in second language learning were rated higher (Lee et al., 2007). Other research showed that in presenting both a facial image and a voice of computer agents, consistency mattered in trust perception (Gong & Nass, 2007). Participants trusted human faces matched with human voices, and computer-generated faces matched with computerized voices more than unmatched cases in an intimate self-disclosure interview by the computer agent.

Although the aforementioned findings significantly increased our knowledge about facilitators of human-computer trust, because the research has been performed on varied types of task situations (e.g., book reviews, competitive game, language learning, interview) and each type of task was a fixed factor for most single studies, we do not know much about the effect of task types on human-computer trust. Thórisson et al. (2016) argue the concept of task is at the core of AI since tasks are used to train and evaluate AI-systems. A task is assigned to an agent as a problem, which includes the way in which the problem is communicated to the agent (Thórisson et al., 2016). Among various ways to categorize different types of tasks an AI-agent can perform and assist humans, there are achievement vs. maintenance tasks (Thórisson et al., 2016), functional vs. social tasks (Gaudiello et al., 2016), simple vs. complex tasks (Guo et al., 2020), and physical vs. virtual service and tangible vs. intangible actions (Wirtz et al., 2018).

Task characteristics, though rarely studied, moderate the relationship between AI-enabled systems and the user (Hancock et al., 2011; Rzepka & Berger, 2018). Users' intention to use an AI-system depends on the fit between technology and task characteristics, and task features significantly affect the decision of technology adoption (Chang, 2010; Daft & Lengel, 1986). Hoffmann and Krämer (2013) found that robots are perceived more favorably in a task-oriented scenario, while virtual characters are preferred in persuasive-conversational contexts. Similarly, Gaudiello et al. (2016) showed users' trust toward robots was higher for functional tasks than social ones, which suggests a better fit between robot technology and functional tasks.

The current study proposes to start building a theory of IVA task based on an existing parameter of tasks Siri can perform: mainly, information search, math calculation, email checking/reading, scheduling, and exchanging humorous conversations with its users. Siri and other IVAs are unique in that "they are both intelligent and social, programmed to follow human communication norms" (Guzman, 2017, p. 3). They perform human-like actions while focused on execution of the different types of tasks based on pre-programmed functions (Skalski & Tamborini, 2007). However, Siri is considered a weak-AI-system, designed to perform only a narrow range of chores; thus, tasks that require expertise, critical thinking, or physical assistance in problem solution are not currently applicable. Hence, a few aforementioned categorizations of tasks such as complex vs. simple, or physical vs. virtual are irrelevant.

Based on previous studies' categorizations of tasks performed by Alagents including robots (Gaudiello et al., 2016; Hoffmann & Krämer, 2013), we categorize a series of tasks Siri can perform or assist humans as functional versus social tasks. Tasks such as information search and math calculation that require cognitive ability can be considered as

functional. Exchanging humorous conversations and supportive messages that involve emotions and empathy can be considered social tasks. Based on this categorization, we explore how human-computer trust in its various dimensions are associated with different types of tasks: mainly, whether the cognitive dimensions of trust (i.e., whether users perceive computer-agents as reliable, understandable, and technically competent; Madsen & Gregor, 2000) are related to functional task and affective dimensions (e.g., users feel personal attachment and faith) to social task performance. The results of the study can increase knowledge about the fit between AI-agents and their task performance and provide a refined theorization of the relationships between human-computer trust and IVA task. Thus, the following hypotheses are proposed:

H1. Human-computer (a) trust in cognitive dimensions will be higher for functional tasks, and (b) trust in affective dimensions will be higher for social tasks assisted by Siri.

## 2.2. Gender effect

How technology's simulated gender matches with cultural expectations within a given context impacts users' perceptions of and interactions with the AI-agents (Damen & Toh, 2019; Nass & Brave, 2005; Nomura, 2017; Guzman & Lewis, 2019). Humans recognize the gender of a machine very easily, and usually perceive female voices as helpful to problem solving and male voices as authoritarian when providing answers (Nass & Brave, 2005). In Lee et al.'s (2000) experiment of computer-generated speech, participants rated the male-voiced computer as more socially attractive and trustworthy than the female-voiced computer, which led them to follow male-voiced computer's suggestions more on social dilemma situations. Mitchell et al. (2011) indicated synthesized female voices were perceived warmer than male voices. McDonnell and Baxter's (2019) experiment revealed that female chatbots' communication was perceived as more affable, friendly, and polite compared to male or gender-neutral chatbots'. Based on the CASA perspective, the authors cautioned users could attribute the success of a system based on their gender-stereotypical perceptions of the AI-agent rather than its functional performances.

However, another recent study found gender of an IVA was not as influential as characteristics of language usage on customer satisfaction and evaluations of hedonic and pragmatic quality (Habler et al., 2019). In their study, submissive and low-status language by an IVA brought a higher-level satisfaction and evaluation; thus, Habler et al. recommended not combining submissive languages with female voices in order to avoid reinforcing gender stereotypes. This aligns with Guzman's (2017) critiques of how Siri's role as a "female secretary" reflects the society's gender stereotypes and reinforces the stereotypes through the inherent power imbalance between users and Siri. Based on a comprehensive review of gender effects in human-robot interaction, Nomura (2017) also questioned whether gendering of robots is really necessary in facilitating the interactions between humans and robots.

Gender has been a significant component in the design of AI-agents and people's interactions with them (Eyssel & Hegel, 2012; Guzman, 2017; Ruijten et al., 2015). Stereotypically, women are perceived as better at relational and emotional tasks versus men at functional and cognitive tasks (Derntl et al., 2010). Nass et al. (1997) earlier found that a male-voiced computer was seen as more influential, and a female-voiced computer was seen as a better teacher of love and relationships. McDonnell and Baxter (2019) identified user satisfaction was lower on female- than male-chatbots when they performed mechanics tasks. There was no difference in satisfaction when they performed banking tasks, supposedly a gender-neutral task. In Damen and Toh's (2019) experiment on a web-based smart lock simulation, exploring the role of AI-agents' location and gendered voice, users tended to trust the automated system more explicitly when it displayed stereotype congruent (female voice-home and male voice-office) design characteristics compared to stereotype incongruent systems (male voice-home and female voice-office). Based on these existing findings,

the following interaction hypothesis is proposed:

H2. Trust toward Siri will be a product of interaction between task type assisted by Siri and Siri's gendered voice.

If such gender stereotypes apply to the participants' evaluations of human-computer trust by distinctive types of task performance, their own gender may also play a role depending on whether their gender matches with the enacted gender of Siri by its voice (Nomura, 2017). Lee et al. (2000) identified cross-over interactions where people found computer-generated speeches of their own gender more attractive and trustworthy. Male participants reported higher conformity, social attractiveness, and trustworthiness toward male-voiced computers and female participants toward female-voiced computers in attractiveness and trustworthiness, which was explained as a social identification process (Lee et al., 2000). More recently, Payne et al. (2013) found that female users preferred three-dimensional, realistic female IVAs in self-service checkout contexts. Ruijten et al. (2015) additionally found that female participants were more susceptible to AI-agents' social feedback compared to males in that they were more likely to change behaviors on an energy saving task.

Contrary to the social identification process that suggests samegender preference, some research shows cross-gender effects that male participants rated female robots as more credible and trustworthy than male robots, while female participants rated the opposite (Siegel et al., 2009). Alexander and colleagues' (2014) experiment involved solving puzzles with a doll robot of which gender was manipulated by its voice and name. Participants felt more comfort with the robot of their opposite gender compared to the one matching their own gender. Thus, research evidence suggests not only the IVA's gender matters, but also the users' and their gender match may impact the level of trust toward the IVA. Whether participants trust same-gender or opposite-gender IVAs more seems to vary by the type of task they perform (e.g., puzzle, navigation, sorting), but existing findings do not clearly suggest a theoretical direction on this issue other than the predictions about preference based on gender stereotypes (see H2). Therefore, we hypothesize an interaction effect of gender match without specifying directions for cross-over or cross-gender effects.

H3. Siri's gendered voice and participant gender match with Siri's will interact to impact the level of trust toward Siri.

## 2.3. Social Presence

Social presence is defined as "a psychological state in which virtual social actors are experienced as actual social actors in either sensory or non-sensory ways" (Lee, 2004, p. 45). Amidst various definitions in the literature (Biocca & Harms, 2002; Nowak & Biocca, 2003), the aforementioned definition of social presence considers the interacting objects broadly as they can include both para-authentic and artificial social actors such as AI-agents. Initially, face-to-face interactions, as the richest communication mode (Daft & Lengel, 1986), were deemed to have the highest level of social presence since interactants had ample opportunities to communicate social cues such as facial expressions, gestures, and voices (Short et al., 1976). However, social presence was not always matched with the bandwidth capacity of a medium; message interactivity and communicators' fluency with technologies could facilitate immediate, intimate, and involved interactions even through lean media such as mobile text or email (Lee, 2004).

Biocca et al. (2003) proposed three dimensions of social presence: a) copresence, referring to not feeling alone and mutual awareness; b) psychological involvement, indicating mutual attention, empathy, and understanding; and c) behavioral engagement, encompassing behavioral interaction, mutual assistance, and dependent action. Based on the first two dimensions, Lee et al. (2005) found social presence mediated participants' social responses toward a robot dog, AIBO. Specifically, as AIBO showed more developmental capacity, people felt a higher level of copresence and psychological involvement with AIBO, which facilitated their social and physical attraction toward AIBO.

Other research also showed that perceptions of social presence served a mediating role, influencing several technology-related perceptions such as usefulness (Karahanna & Straub, 1999), enjoyment (Hassanein & Head, 2006), involvement and arousal (Fortin & Dholakia, 2005), and trust (Gefen & Straub, 2004). Hess et al. (2009) showcased how social presence in the context of a website interface facilitated users' trust toward recommendation agents of the website. Gefen and Straub (2004) provided three reasons for the relationship between social presence and trust based on the delivery of social cues. First, it is easier to hide and engage in an untrustworthy activity when fewer social cues are conveyed. Media that convey more social cues presence thus can be viewed as more transparent, and consequently, more trustworthy. Second, trust develops when a trustor can expect the trustee's behaviors. Increased social presence—because it conveys more information—may lead to trustors' beliefs that such assessments are easier, thereby enhancing trust. Third, trust may increase when the trustor perceives more investment in the relationship and concludes that the trustee has more to lose from untrustworthy behavior (Gefen & Straub, 2004).

Although logical, most applications of social presence concept rely on the assumption of a linear relationship between media richness and social presence. This assumption ignores the fact that social presence is an individual's psychological state, not necessarily determined solely by a medium's capability (Biocca et al., 2003; Lee & Nass, 2004). Richer media or embodied AI-agents such as robots can facilitate a higher level of social presence compared to leaner media or IVAs, but not always. Message content, interactivity of communication, personal experiences and proficiency in communicating with AI-agents can all variably impact social presence.

Guzman and Lewis (2019) suggest communication scholars of AI look beyond addressing HMC against existing theories of human-to-human communication to critically engage in questioning whether the interaction can be understood by the existing categories, and if not, what new classifications can be developed or old boundaries redrawn. Therefore, even if social presence theory is easily applicable to various types of HMC, including the interaction with AI-agents, exactly how communicators' feeling of social presence lead to trust in different types of AI-systems, if at all, needs to be empirically investigated. For example, when interacting with an IVA similar to Siri, can a user still feel the high level of social presence as when interacting with embodied agents, like humans or social robots, especially if they are familiar with using Siri (although interaction with Siri is based only on audio or textual cues)? How would different gender of Siri's voice, its match with a user's gender, and distinct type of task impact social presence with Siri? The current study proposes the following two research questions in lieu of the above discussion about social presence and its potential association with trust of an AI-agent:

RQ1. How do (a) type of task assisted by Siri, (b) gender of Siri's voice, and (c) gender match with the participant relate to social presence felt during interactions with Siri?

RQ2. How does social presence during interactions with Siri relate to trust toward Siri?

# 2.4. Relationships with AI-agents

Finally, the current study explores how social presence and trust toward Siri is associated with humans' comfort level toward Siri in various relational scenarios: as a coworker, supervisor, or friend. We believe examining the relationships among these constructs (i.e., social presence, trust, and comfort toward Siri) is important to understand which components of HMC may facilitate more positive or negative attitudes and an (un)comfortable feeling toward an IVA. Higher level of social presence and trust may generally bring more favorable attitudes toward Siri or vice versa, but few empirical studies examined how such associations are formed in different types of relational scenarios (e.g., coworker, supervisor, and friend).

IVAs and other AI-agents can be utilized in a formal employment

setting, thus considered as coworkers or even supervisors if employees must follow directions from the AI-agents (Groom & Nass, 2007). Although Groom and Nass (2007) argued lack of human-like mental model and sense-of-self would make robots untrustworthy as "teammates" in a workgroup, robots and IVAs that can interact with humans with speech and potentially read emotions with the advanced AI-systems may still serve as companions (Ho et al., 2018). Robotic pets and therapeutic robots are currently available, and humans develop surprisingly quick emotional attachment toward these human alternatives (Turkle, 2012). As IVA users' current attitudes likely influence their future attitudes and behaviors related to similar or more advanced types of technologies (Guzman & Lewis, 2019), exploring the facilitator of comfortable attitude toward AI-agents as various relational partners are worthwhile. Thus, we ask the following:

RQ3. How do social presence and trust toward Siri relate to comfort level with Siri as a (a) coworker, (b) supervisor, or (c) friend?

#### 3. Methods

This study utilized an in-lab experiment where participants were invited to interact with Siri via their own iPhone and answer questions provided in an online survey, hosted on Qualtrics. Undergraduate students in a research pool of a large West South-Central university in the United States were recruited for the experiment and those who participated in the study received extra credit for their enrolled courses of choice. Data was collected from mid-October 2018 till late April 2019 and a total of 163 participants' responses out of 170 attempted trials were used for data analysis. Seven responses were removed initially from the dataset because some of them were not iPhone users, thus not eligible to participate, and others had to leave in the middle of the experiment due to personal situations, nullifying their participation in the study.

# 3.1. Experimental Setting

Once participants agreed to the informed consent and were identified as a Siri user, depending on random assignment, they were asked to change Siri's voice setting into either American female (n = 86) or male (n = 77). If the participant's iPhone was already configured to the condition they were assigned, no change was necessary. Participants were also randomly assigned to a task-condition (functional,  $n=82\ vs.$ social, n=81). Participants gender was either matched (n=90) or unmatched (n = 73) with the Siri's voice. Approximately even numbers of participants were assigned in each setting  $(2 \times 2 \times 2)$  as a result of randomization. Functional tasks of the experiment involved getting information about weather, stock market, and locations from Siri and asking Siri to perform mathematical calculations. Social tasks involved asking humorous questions to Siri, such as "How much do you earn?" "Siri, do you sleep?" or "Will pigs fly?", and getting to know Siri, such as "What does Siri mean?" "Why did Apple make Siri?" or "How old are you, Siri?" Participants were asked to fill in Siri's answers to a total of 15 pre-determined questions presented in the same order for everyone from the online survey in either the functional or social task setting. Participants asked those questions to Siri first, and then entered what they heard as Siri's responses in the survey as a way for the researchers to verify they had actually performed the tasks. Due to having this record, we did not include a manipulation check for the task type. For a full list of questions used for each type of task, please see the Appendix I.

However, we asked an open-ended question verifying the gender of Siri after the interaction and 11 out of 163 participants answered inconsistent to the gender they were assigned during the experiment. Two out of 11 participants answered they did not think Siri has a gender and one answered they had Australian male when all three of them were assigned to the American female voice setting. Since the wording of our question was, "What was the gender of your Siri?" we think some participants may have misunderstood our intent and thought that the

question was about their previous setting. One participant answered, "Before this study my Siri was set to American female" and another said, "American male. I usually have American female." Thus, we do not claim our manipulation check for the Siri's gender was 100% successful; but considering the misunderstanding of the question evidenced by the participants' answers, we believe over 93% of the participants identified the gender of Siri's voice as intended by the experimental design. When being randomly assigned to a different gender (and task group), we also showed a picture of female (for female voice) or male (for male voice) to aid with recall of the gender and to make sure they were changing their setting to the assigned one correctly.

#### 3.2. Sample

There were more female (n = 109, 66.9%) than male participants, and participants' age ranged from 18 to 27 (M = 19.6, SD = 1.5). Participants identified their race as white (n = 115, 70.6%), Asian (n = 22, 13.5%), non-white Hispanic (n = 9, 5.5%), African/Black (n = 6, 3.7%), and others including Native American (n = 11, 6.7%). Over 40% of participants were first-year students (n = 67), 32.1% were sophomores (n = 50), 14.1% juniors (n = 22), and 10.4% were seniors (n = 17). Seven students did not identify their years in college. The majority of participants were US citizens (n = 141, 86.5%), and the rest were international students (12.9%) $^1$ .

#### 3.3. Measurements

#### 3.3.1. Trust

Participants' trust toward Siri was measured using a 25-items scale developed by Madsen and Gregor (2000). The questions included the following five constructs with five measurement items per each dimension (5-point Likert-type, 1 = strongly disagree, 5 = strongly agree): perceived reliability, perceived technical competence, perceived understandability, faith, and personal attachment. A confirmatory factor analysis (CFA) was performed to assess validity of the scale as Madsen and Gregor suggested some dimensions were unstable from their initial analysis. An initial model fit from CFA was:  $\chi^2$  (265) = 409.91, p < .001, Steiger-Lind root mean square error of approximation (RMSEA) = 0.06 with 90% CI = [.05, .07], Bentler comparative fit index (CFI) = 0.92, and Standardized root mean square residual (SRMR) = 0.06 (Hu & Bentler, 1999; Wang & Wang, 2012).<sup>2</sup> After allowing for error covariances between the latent constructs and error covariances of items within subdimensions when suggested via modification indices for six cases, another CFA was performed and a better fit was found:  $\chi^2$  (259) = 317.13, p = .008, RMSEA = 0.04 with 90% CI = [.02, .05], CFI = 0.97, and SRMR = 0.05. Though the chi-square value was significant, the relative chi-square ( $\chi^2$ /df) was reasonable with 1.22 (Tabachnick & Fidell, 2013, recommend < 2). Based on the CFA results, five dimensions of the trust scale were calculated as separate average scores, and all had acceptable range of scale reliability: perceived reliability (M = 3.41, SD = 0.73,  $\alpha$  = .79), perceived technical competence (M = 3.58, SD = 0.72,  $\alpha = .74$ ), perceived understandability (M = 3.82, SD = 0.80,  $\alpha = .83$ ),

faith (M = 2.53, SD = 0.89,  $\alpha$  = .86), and personal attachment (M = 2.13, SD = .080,  $\alpha$  = .86).

#### 3.3.2. Social presence

A modified version of Biocca et al.'s (2003) social presence scale was used to measure two dimensions: copresence and psychological involvement. The third dimension of behavioral engagement was not relevant to the context of the current study as it involves eye contact or nonverbal mirroring behaviors of interactants. Copresence was measured by the following three questions: (a) How much did you feel as if you were interacting with an intelligent being? (b) How much did you feel as if you were alone? (reverse-coded) (c) How much did you feel as if you and Siri were communicating with each other? Psychological involvement was measured by the following two questions: (a) How much attention did you pay to Siri? (b) How much did you feel involved with Siri? Responses were measured with a 10-point scale (1 = not at all, 10 = very much). A scale reliability test suggested eliminating the reverse-coded item would increase the internal consistency of the copresence dimension. Both dimensions of social presence were calculated as aggregated scores: copresence (M = 9.89, SD = 3.91, r = .54) and psychological involvement (M = 10.86, SD = 4.41, r = .52).

#### 3.3.3. Attitudes toward Siri

Three questions were asked about participants' comfort level toward Siri as a (a) supervisor, (b) coworker, and (c) friend. The questions were worded as "How would you feel if Siri was your...?" and the responses were measured with a 7-point Likert type scale (1 = extremely uncomfortable, 7 = extremely comfortable). Participants showed a slightly more favorable attitude toward Siri as their coworker (M = 3.58, SD = 1.45) than as a supervisor (M = 3.06, SD = 1.38), t(162) = 6.13, p < .001, or as a friend (M = 2.99, SD = 1.52), t(162) = 6.48, p < .001.

#### 3.3.4. Controls

Participants also reported their length of Siri usage, and more than 40% had used it over four years. Less than 10% had been using it less than one year, and 8.6% between one to two years. Almost 20% of participants had been using Siri between two to three, and between three to four years each. The proficiency of Siri usage was measured with a 5-point scale (1 = Novice, 5 = Expert; M = 3.08, SD = 0.78). These variables were considered as covariates or controls for the data analyses.

## 3.4. Analytical Strategy

To test the between-subject effects of Siri's gendered voice (female vs. male) with two different types of tasks (functional vs. social) and gender match (matched vs. unmatched) with participants' in main effects of task types (H1s) and their interactions (H2 & H3) on the level of trust, a multivariate analysis of covariance (MANCOVA) with a full-factorial design was performed with the proficiency of using Siri and length of usage as covariates. To answer RQ1s about the effect of task, gender, and gender match on social presence, another MANCOVA was performed. Hierarchical regressions were used to answer RQ2 and 3 concerned with the relationships among social presence, trust, and comfort toward Siri.

# 4. Results

#### 4.1. Results for H1s, H2, and H3

A MANCOVA was conducted to test the main effects of task type (H1s) alongside an interaction between Siri's gendered voice and task type (H2) and another interaction between participants' gender and Siri's gendered voice (H3), while controlling for the participant's proficiency and length of Siri usage. The five dimensions of human-computer trust were used as dependent variables. Box's test of equality of covariance matrices was significant at p < .01 level; thus,

<sup>&</sup>lt;sup>1</sup> Due to the potential impact of English fluency on participants' interactions with Siri, we compared all major variables including social presence, trust, and comfort level with Siri between international students group and US citizens group. No variable was significantly different between the two groups in the independent sample t-tests.

 $<sup>^2</sup>$  Kline (2011) recommends reporting measures of incremental fit (i.e., comparative fit index [CFI]), parsimony (i.e., root mean square error of approximation [RMSEA]), and absolute fit (i.e., standardized root mean square residual [SRMR]). Low RMSEA and SRMR values, near 0, are preferable; values of CFI near 1.0 indicate the best fit (Kline, 2011). Hu and Bentler (1999) provide more specific values for each type of fit: CFI  $\geq$  .95, RMSEA  $\leq$  .06, and SRMR  $\leq$  .08.

Pillai's Trace for multivariate test was used for robust interpretation (Tabachnick & Fidell, 2013). Levene's test of equality of error variances was insignificant for any of the dependent variables. Therefore, the error variances of the populations being compared were considered consistent.

Pillai's Trace for multivariate test was not significant for any type of interaction effects. However, Pillai's Trace (.08) was significant for the main effect of task, suggesting varying levels of trust by task type, F (5, 149) = 2.42, p < .05, partial  $\eta^2 = .08.^3$  Between-subject effects were significant for all but the perceived reliability dimension of trust. There was no significant main effect of Siri's gendered voice, or gender match between the participant's and Siri's. Still, proficiency in Siri usage showed a significant main effect on trust, Pillai's Trace = .13, F (5, 149) = 4.33, p < .01, partial  $\eta^2 = .13$ .

Subsequent analysis of univariate tests revealed that in three out of five dimensions of trust, the nature of task made significant differences. Particularly, perceived reliability, F (1, 151) = 4.77, p < .05, partial  $\eta^2$  = .03, technical competence, F (1, 151) = 7.36, p < .01, partial  $\eta^2$  = .05, and perceived understandability, F (1, 151) = 8.89, p < .01, partial  $\eta^2$  = .06, were reported higher when Siri assisted with functional tasks compared to social tasks. These three dimensions of trust were all considered as cognitive aspects of human-computer trust (Madsen & Gregor, 2000); thus, the association between functional tasks and cognitive trust (H1a) was supported. However, faith and personal attachment, affective dimensions of human-computer trust, did not vary by the task type, which contradicted H1b. Figure 1 visualized these results in a bar chart.

A further examination of between-subject effects revealed one significant interaction between participants' gender and Siri's gendered voice over one factor of trust, faith, F (1, 153) = 4.30, p < .05, partial  $\eta^2$ = .03. Participants who were women and interacted with Siri's female voice reported lower levels of faith in Siri (M = 2.53, SD = .79) than men who interacted with a Siri's male voice (M = 2.83, SD = .92). Furthermore, women who interacted with a male Siri's voice reported lower levels of faith (M = 2.30, SD = .89) than men who interacted with a female's voice (M = 2.72, SD = 1.02). According to these results, both gender participants seemed to trust Siri's voice matching with their own gender slightly more, and woman participants trusted Siri in both female's and male's voices less than the men (see Figure 2 for the interaction graph). In sum, the interaction between Siri's gendered voice and task type (H2) was unsupported, but the interaction between participants' gender and Siri's gendered voice (H3) was partially supported in one dimension of trust, faith.

#### 4.2. Results for RQ1 though RQ3

#### 4.2.1. Results for RQ1

The effect of Siri's (a) task type, (b) gendered voice, and (c) gender match with participants on social presence was examined with another MANCOVA with two dimensions of social presence (i.e., copresence and psychological involvement) as dependent variables. The proficiency and the length of Siri usage were covariates. Box's test of equality of covariance matrices was insignificant; and Levene's test of equality of error variances was also insignificant for any of the dependent variables.

The results of MANCOVA showed no significant difference in social presence based on the task type, gender of Siri's voice, or gender match with participants. No significant interaction was observed. Only the level of proficiency had a statistically meaningful relationship with social presence, F (2, 139) = 5.59, p < .01, partial  $\eta^2$  = .07. Both copresence, F (1, 149) = 10.55, p < .01, partial  $\eta^2$  = .07, and psychological

involvement, F (1, 149) = 5.28, p < .05, partial  $\eta^2$  = .04, had positive relationships with the level of proficiency, meaning highly proficient Siri users reported they felt more copresent and psychologically involved during the interaction with Siri compared to novice or average users.

#### 4.2.2. Results for RQ2

A series of hierarchical regression was conducted to examine the relationships between social presence and the five dimensions of trust toward Siri. After considering individuals' demographics (i.e., age, gender, race), proficiency and length of Siri usage, and gender of Siri's voice and task type, copresence and psychological involvement were entered as the last block of predictors in the regressions. Table 1 summarizes the results of the hierarchical regressions on the five dimensions of trust. Due to the limited space, only the final steps of the hierarchical regression results are presented.

First, the dimension of perceived reliability had a positive association with the level of psychological involvement after controlling for the effect of task type (see the second column of Table 1). Participants who performed functional tasks assisted by Siri tended to report higher levels of reliability; but regardless, if they felt highly involved psychologically during the interaction with Siri, they also perceived reliability of Siri higher. Adding the social presence variables at the last step of regression increased 12% of the explained variance in perceived reliability, F-change (2, 153) = 11.80, p < .001.

The dimension of technical competence had positive associations with the level of both copresence and psychological involvement after controlling for the effect of task type and length of Siri usage (see the third column of Table 1). Participants who performed social tasks assisted by Siri and used Siri longer reported lower levels of technical competence of Siri; but, if they felt higher levels of copresence and psychological involvement during the interaction with Siri, they perceived higher levels of technical competence. Adding the social presence variables in the last step explained 11% more variance in technical competence, F-change (2, 153) = 11.08, p < .001.

For the third dimension of trust, perceived understandability, the level of copresence had a positive association after controlling for the effect of task type and proficiency of Siri usage (see the fourth column of Table 1). Participants who performed functional tasks assisted by Siri and were highly proficient in using Siri reported higher levels of understandability of Siri; but regardless, if they felt higher levels of Siri's copresence during the interaction, they could understand how Siri worked better. Adding the social presence variables at the end of the regression increased the amount of explained variance in perceived understandability by 5%, F-change (2, 153) = 4.56, p < .05.

The faith dimension of trust was examined as the next criterion variable and the level of copresence had a positive association after controlling for the effect of task type and participants' gender (see the fifth column of Table 1). Participants who performed functional tasks assisted by Siri and identified as men reported higher levels of faith in Siri; additionally, if they felt Siri was highly copresent during the interaction, they reported even more faith in Siri. Adding the social presence variables as the last block of the regression brought a 12% increase in the amount of explained variance in faith, F-change (2, 151)  $=12.07,\,p<.001.$ 

The last dimension of trust, personal attachment was examined through the same set of predictors in the regression. The results showed that only copresence was positively associated with personal attachment toward Siri, after controlling for the effect of proficiency level. Participants with higher proficiency of using Siri reported higher levels of attachment toward Siri; further, if they felt higher levels of Siri's copresence during the interaction, they also reported higher levels of personal attachment. Adding the social presence variables as the last block explained 12% more variance in personal attachment, F-change (2, 153) = 12.66, p < .001.

Based on the above results, the level of social presence participants felt during interactions with Siri seemed to be closely associated with the

 $<sup>^3</sup>$  A post hoc power analysis was conducted using G\*Power 3.1.9.4 (Faul et al., 2007) to test the difference across eight groups (2x2x2) using F tests with a medium effect size (f = 0.30) and an alpha of .05. Result showed that the sample size of 163 achieved a power of .84.

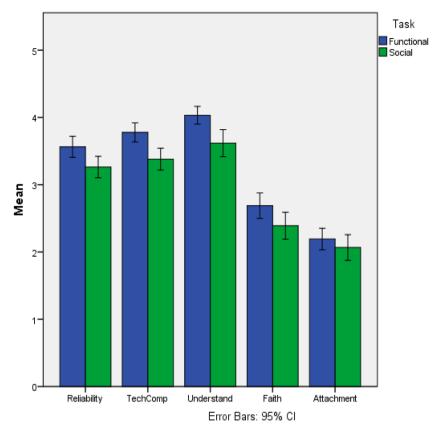
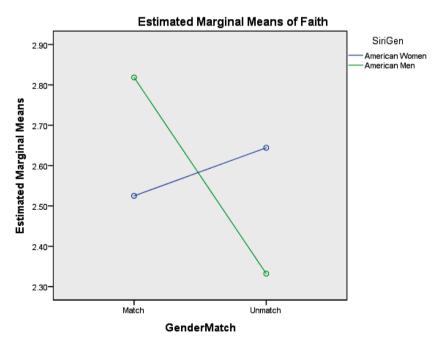


Figure 1. Differences in human-computer trust of five dimensions by task types.



Covariates appearing in the model are evaluated at the following values: Approximately, How long have you been using Siri? = 3.79, How proficient are you in using Siri? [Recoded] = 3.08

Figure 2. Interaction between participants' gender and gender of Siri's voice on Faith

level of trust they reported about Siri. Especially, feelings of copresence with Siri were significantly related to four out of five dimensions of trust, and psychological involvement with Siri was related to two out of five dimensions of trust (see Table 1).

#### 4.2.3. Results for RQ3

Finally, three hierarchical regressions were performed to examine the relationships among social presence, trust, and comfort level with Siri as a (a) coworker, (b) supervisor, and (c) friend. First, participants'

**Table 1**Trust on Siri predicted by social presence

	Reliability β (t)	TechComp β (t)	Understand $\beta$ (t)	Faith β (t)	Attachment β (t)
Age	-0.08 (-0.96)	-0.11 (-1.47)	-0.02 (-0.19)	0.00 (0.03)	0.10 (1.31)
Gender (female $= 1$ )	0.03 (0.42)	0.00 (0.03)	-0.07 (-0.89)	0.12 (1.55)*	0.04 (0.58)
Race (white $= 1$ )	0.06 (0.77)	-0.02 (-0.27)	0.00 (-0.05)	0.07 (0.92)	0.04 (0.53)
Length of Usage	-0.11 (-1.37)	-0.17 (-2.25)*	-0.03 (-0.37)	-0.03 (-0.40)	-0.02 (-0.29)
Proficiency	0.05 (0.51)	0.03 (0.31)	0.15 (1.76)*	0.09 (1.11)	0.26 (3.34)**
Siri Gender ( $f = 1$ )	-0.01 (-0.07)	-0.08 (-1.05)	0.02 (0.29)	-0.03 (-0.36)	0.01 (0.12)
Task (functional $= 1$ )	-0.19 (-2.51)*	-0.25 (-3.35)**	-0.23 (2.99)**	-0.16 (-2.19)*	-0.08 (-1.11)
Copresence	0.18 (1.84)	0.20 (2.16)*	0.22 (2.31)*	0.35 (3.85)***	0.38 (4.19)***
Psych Involvement	0.25 (2.66)**	0.23 (2.50)*	0.07 (0.74)	0.08 (0.94)	-0.01 (-0.12)
Adjusted R <sup>2</sup>	.16***	.19***	.12***	.19***	.22***

p < .05\*\* p < .01

attitude toward Siri as a coworker was significantly and positively associated with the level of copresence and faith, after controlling for the effect of their age, and the three variables explained about 30% of variance in the comfort level toward Siri. Thus, it seems higher levels of copresence and faith are associated with a more comfortable feeling toward Siri as a coworker (see the second column of Table 2). Older participants were feeling more comfortable with Siri as a coworker. Perceived understandability of Siri was positively associated with the comfort level, but the relationship was only marginally significant according to the conventional criteria.

Participants' comfort level with Siri as a supervisor was also significantly associated with the faith dimension of trust. Although the level of copresence was significantly related to the attitude in step 3 of the regression, once the five dimensions of trust were entered in the last step, copresence was no longer a significant predictor of comfort (see the third column of Table 2). Adding the five trust variables during the last step of regression increased the amount of explained variance by 9.5%, F (5, 148) = 4.43, p < .01. The result seemed to suggest that a higher level of faith in Siri facilitated a more comfortable feeling toward Siri as a supervisor.

For the attitude toward Siri as a friend, copresence was still significantly associated, but a different dimension of trust, personal

 Table 2

 Regressions on comfort toward Siri as various relational partners

	Coworker	Supervisor	Friend
	β (t)	β (t)	β (t)
Age	0.21	0.06 (0.70)	0.17 (2.26)*
	(2.72)**		
Gender (female = baseline)	-0.03 (-0.35)	-0.02 (-0.20)	0.01 (0.13)
Race (white = baseline)	-0.04 (-0.58)	-0.02 (-0.26)	-0.03 (-0.45)
Length of usage	-0.05 (-0.66)	-0.004	-0.01 (-0.14)
		(-0.05)	
Proficiency	-0.10 (-1.16)	-0.03 (-0.37)	-0.17 (-2.03)
			*
Siri Gender (female = baseline)	-0.04 (-0.48)	-0.04 (-0.41)	-0.19 (-2.47)
			*
Task (functional = baseline)	0.04 (0.54)	0.13 (1.56)	0.16 (2.22)*
Gender Match (matched = baseline)	-0.09 (-1.13)	-0.01 (-0.12)	0.002 (0.02)
Copresence	0.26	0.14 (1.27)	0.33
	(2.69)**		(3.46)**
Psychological Involvement	0.09 (1.03)	-0.001	0.02 (0.25)
		(-0.01)	
Reliability	0.05 (0.49)	-0.07 (-0.58)	-0.02 (-0.15)
Technical Competence	-0.17 (-1.42)	-0.08 (-0.61)	-0.14 (-1.24)
Understandability	$0.19 (1.93)^{\dagger}$	0.11 (1.01)	0.10 (0.98)
Faith	0.23 (2.18)*	0.32 (2.72)**	0.14 (1.36)
Personal Attachment	0.07 (0.75)	0.13 (1.28)	0.26
			(2.86)**
Adjusted R <sup>2</sup>	.297**	.135**	.317**

p < .10

attachment was related to the attitude. The six variables while controlling for the effect of participants' age, proficiency of Siri usage, Siri's gendered voice, and task explained 31.7% of the variance in comfort level with Siri as a friend. Considering the result, a higher level of Siri's copresence and personal attachment toward Siri felt during the interaction seem to facilitate a more comfortable feeling toward Siri as a friend (see the fourth column of Table 2). Interestingly, participants were feeling more comfortable with Siri as a friend when the voice was a female's and they engaged in social tasks. Also, notably, younger and more proficient Siri users were less comfortable with Siri as a friend.

#### 5. Discussion

The current study used an experiment to test how gender of Siri's voice, type of task assisted by Siri, and participants' gender match with Siri's influenced their trust and social presence felt toward Siri during the interaction. In addition, participants' comfort level with Siri as a coworker, supervisor, or friend was explored vis-à-vis their social presence and trust felt during the human-machine communication. As voicebased AI-technologies such as Siri, Alexa, and Cortana are being rapidly adopted by consumers, and soon other types of AI-operated systems such as humanoid robots may be utilized frequently in daily lives, how technology users interact with such AI-agents and form trust in them is worthy of scholarly attention (Guzman & Lewis, 2019). Findings of the current study seemed to partially confirm previous theory and research based on the computers-are-social-actors (CASA) perspective (Reeves & Nass, 1996; Lee & Nass, 2010) but challenge the stereotypic effect of (enacted) gender of the machine. Also, our findings revealed distinct associations among social presence, trust, and participants' comfort with Siri. We discuss in more detail about the implications and contribution of these study results in the following.

First, the study hypothesized a difference in participants' trust of Siri between functional and social tasks assisted by Siri. In three (i.e., perceived reliability, technical competence, perceived understandability) out of five dimensions of trust measured in this study, we found statistically meaningful differences with participants reporting higher trust when engaged in functional tasks with Siri. This finding indicated humans could rely on AI-machines such as iPhones with Siri when conducting functional tasks such as searching for information related to weather, the stock market, or performing math calculations, and they perceived the machine reliable, competent, and understandable. These three dimensions of trust were theorized as the cognitive components of human-computer trust (Madsen & Gregor, 2000), and the association between functional task performance and cognitive dimensions of trust seemed reasonable. However, social tasks such as exchanging humorous conversations with Siri did not seem to facilitate participants' affective trust of the IVA such as personal attachment or faith. Participants might have perceived social tasks as still uniquely human and/or Siri's answers to the non-sensical questions (for having humorous conversations) as awkward.

 $<sup>^{**}</sup>$  p < .001. TechComp: Technical competence; Psych Involvement: Psychological involvement.

<sup>\*</sup> p < .05

<sup>\*\*</sup> p < .01.

Second, the study did not find an interaction effect between Siri's gendered voice and the task type, contrary to some previous research findings (e.g., Damen & Toh, 2019; Derntl et al., 2010). The reason for this insignificant interaction may have to do with the nature of our task assisted by Siri. Based on the existing literature, we expected participants might trust a male's voice more when performing functional tasks and a female's voice with social tasks. Previous research showed female chatbots were trusted less when performing mechanics tasks compared to male chatbots, and female-voiced computers were thought of as better teachers of love and relationships. Thus, gender stereotypes seemed to operate in users' interactions with computers and AI-agents. However, our task categorization of functional vs. social might not have caused gender stereotyping in participants' cognitive process while interacting with gendered voices of Siri. Another possibility is that young college student participants in our sample might not have developed strong gender stereotypes, so they did not associate a particular type of task with a gendered voice of Siri.

Although the expected interaction effect between the task type and gendered voice was not found in this study, there was one interaction identified between Siri's gender and participants' gender on the faith dimension of trust. Participants of both genders reported slightly higher levels of faith in their same gender voice of Siri, which could be understood as a social identification process (Lee et al., 2000; Payne et al., 2013) and women reported much lower levels of faith in both male- and female-voiced Siri compared to men. Faith dimension of trust involved statements such as "I believe advice from Siri even when I don't know for certain that it is correct" and "When Siri gives unusual advice, I am confident that the advice is correct." This gender difference in the level of faith toward Siri aligns with many previous studies showing men tend to have higher efficacy, knowledge, experiences, and confidence in technology acceptance and usage (Chou et al., 2011; Lin et al., 2012). A meta-analysis of 50 articles between 1997 and 2014 reported men tend to have more positive attitudes toward technology than women; the attitude gap is reducing in dimensions like affect and self-efficacy, but not in the dimension of belief (Cai et al., 2017). Considering the measurement items of faith dimension of this study, 'belief' and 'faith' in technology seem to overlap and our finding of men participants' stronger faith in Siri evidences the remaining gender gap in attitude.

Third, the study explored the impact of Siri's gendered-voice, task type, and gender match with participants on their level of social presence felt during the interaction. The analysis did not find any meaningful differences in the two dimensions of social presence (i.e., copresence & psychological involvement) by any factor. This result partially confirmed a previous finding (i.e., Kavya et al., 2019) that even with a different type of social presence measure (i.e., Biocca et al., 2003), no impact of gender or task type was found<sup>4</sup>. These combined results seem to suggest the feeling of being together and being psychologically involved while interacting with an AI-machine is not profoundly influenced by either factor. Research shows that social presence is influenced by interactivity and vividness of messages (Fortin & Dholakia, 2005), and Lee (2004) and Biocca et al. (2003) suggest social presence is a psychological state of an interactant, rather than a technology's characteristic. Thus, a medium with a larger bandwidth and capability of delivering multiple cues does not necessarily facilitate a higher level of social presence, but communicators' familiarity with each other's style, and interactivity and liveliness of conversations have a stronger influence on social presence. The result of this study showing highly proficient Siri users experiencing higher levels of social presence can be understood in such contexts.

Fourth, in line with previous studies (Gefen and Straub, 2004; Hess et al., 2009), we found close connections between the social presence and trust participants felt toward Siri. The more participants felt like they were communicating with Siri and interacting with an intelligent being, the more they perceived Siri as technically competent, understandable, believable, and felt personally attached to Siri. Also, the more participants felt they were involved with Siri and paid attention to their interaction with it, the more strongly they perceived Siri as reliable. This finding may suggest that humans could feel a good amount of social presence [sample mean was 9.89 (range 1 - 18) for copresence and 10.86 (range 0 - 20) for psychological involvement] while interacting with an AI-machine just through voices, which again adds support to the claim that social presence is a psychological state, not necessarily determined solely by a medium's capacity (Biocca et al., 2003; Lee, 2004); and that feeling of presence was meaningfully associated with their willingness to trust the machine. The possibility of feeling an AI-agent's social presence and being able to trust it in its task assistance forms bases of having meaningful relationships with various forms of AI-agents.

Finally, this study examined how the social presence and trust participants experienced during the interaction with Siri were related to their comfort level with Siri when imagined as various types of relational partners (i.e., coworker, supervisor, friend). As a result, copresence was significantly associated with the comfortable attitude toward Siri as a coworker and a friend. Faith dimension of trust was related to a comfortable feeling toward Siri as a coworker and a supervisor whereas personal attachment dimension of trust was related to the comfort toward Siri as a friend. These results seem to suggest that in order for humans to work with an AI-agent as a coworker or even as a supervisor in the near future (Groom & Nass, 2007), maintaining a high level of copresence in communication and faith in the machine will be critical. Just as humans naturally feel attachment to their friends and trust them, similar mechanisms may work when we form a friendship with a communicative AI-agent (Ho et al., 2018)

When an IVA like Siri understands human speech better with its improved natural language processing, as in the ability to have lively and fun conversations with us, a feeling of intense copresence could become even more possible. As copresence was closely related to almost all dimensions of human-computer trust in this study, more development in natural language processing of IVAs would be desirable to increase the level of trust toward AI-agents. It is also notable that cognitive dimensions of trust (i.e., reliability, understandability, and technical competence), associated with functional task performance with Siri, were not related to comfortable feelings toward Siri in varied relational arrangements. The results seem to imply that humans may trust AImachines for getting assistance in functional tasks (e.g., information search, math calculation), but when it comes to forming relationships, affective dimensions of trust (i.e., faith, personal attachment) become a deal-breaker. Research shows chatbot users who engage in more complex tasks prefer turn-based conversational experiences with chatbots that are capable of natural language conversation, such as the ability to perform the human norm of appropriately ending a conversation (Jain et al., 2018). Thus, in order to facilitate affective trust and its consequential comfort toward AI-agents, improving natural language processing of AI-agents seems to be a key design issue.

# 5.1. Limitations and Future Directions

Although this study utilized an experimental approach to test the effect of gender and task on various dependent variables (e.g., social presence, trust, and comfort toward Siri) while controlling for other unknown effects through random assignment, the participant sample was limited both in terms of nature and size. College students, the population of our sample, are the most active users of recent communication technologies; but we did not use random sampling. Thus, the study results have limited generalizability. A future study should

<sup>&</sup>lt;sup>4</sup> In a previous study of an online experiment, parts of Tu's (2002) social presence and privacy questionnaire were adopted to measure two dimensions of social presence: social context and interactivity. Since there were no significant differences of social presence between gender or task type in the previous study, we adopted Biocca et al.'s (2003) measure for the current in-lab study to examine whether the previous result was a measurement issue.

consider random sampling of a larger and broader population of Albased communication system users.

Another limitation to note is how this study could not incorporate the aspect of technology users' preexisting relationships with IVAs including Siri, what they normally use it for, and its prior voice setting. Although we had control variables such as proficiency and length of usage, and proficiency was indeed positively associated with the level of social presence, these two variables may not capture all dynamic aspects of smartphone users' relationships and daily routines with Siri (Gambino et al., 2020). Some users may use Siri quite frequently for various functions and exchange jokes with it whereas other users may not even remember that Siri exists in their phones (or intentionally avoid using it due to privacy concerns). This study did not track whether participants had to switch to another voice setting from their usual setting; if participants had a female's (or a male's) voice as a default, changing it to the other voice due to the experiment may have brought about discomfort and unfamiliarity. There may have been a novelty effect that listening to an unfamiliar voice could increase participants' attention and engagement. Based on our correspondence with the research associates who monitored participants' in-lab sessions, we believe almost all participants had a female's voice as their default and had to switch only when they were assigned to a male's voice group. However, a future study should consider these existing settings and relationships with IVAs, which can be changed over time, and confounding factors when examining users' social presence, trust, and comfort toward AI-agents.

Finally, if the experiment was performed using different types of IVAs or more advanced and difficult sets of tasks, the results might have been different. Siri is a relatively well-known and wide-spread AI-system available in one of the most popular smartphone systems, iPhone, or iOS. Thus, participants may have acquired a certain level of trust and comfort level in using the system before the experiment and if they were given a new, unfamiliar system, the level of social presence, trust, and comfort might have been lower and gender and task effect may operate differently. Also, due to Siri being a relatively weak AI-system, being able to assist only simple tasks, our choice of task selection was limited by the currently available functions Siri can perform. Thus, we cannot claim findings of this study are readily generalizable to other types of IVAs, or more advanced AI-systems that can perform more complicated tasks

# 5.2. Design Implications

Findings of this study can inform future design of IVAs and other types of AI-agents to enhance human capacity at work. If designers of AIsystems value affective dimensions of human-computer trust such as faith and attachment, it is recommended that Siri and other types of IVAs need to be improved in socializing capability. Research shows that some users prefer social-agent orientations (i.e., having small talks and chatting) more than others and such preferences make meaningful differences in their evaluations of AI-agents or chatbots (Liao et al., 2016). Although the impact of individual differences like preference was controlled in our experiment by random assignment (e.g., functional vs. social tasks), future studies can focus on measuring participants' preference for human-like social interactions with IVAs and consider its impact on their trust toward the agents. Depending on the results, future AI-agents may have a feature to customize the level of sociability so that users who enjoy engaging in small talks with IVAs can choose a "relational mode" whereas others who prefer to have a task-oriented IVA without any socializing aspect can choose a "task mode." Perhaps one may prefer both modes combined; thus, having a third option of both task and relational modes on would be considerable.

Furthermore, individuals tend to act in patterns familiar to them. However, with a sharp increase in online and distributed work due to the COVID-19 pandemic, there are new challenges arising for individuals regarding maintaining a balance between work activities and personal responsibilities. In this new climate, IVAs like Siri can help alleviate the

stressors connected with managing routine tasks. The results from this study showcased many individuals felt comfortable with Siri as a coworker. Designers of IVAs can build on this finding and create and market ways for individuals to engage in functional tasks with even greater ease of access. Ease of access refers to the user's accessibility to the IVA with the least amount of effort and resources. With IVAs acting as a buffer and handling routine tasks in an efficient manner, individuals will be able to focus their energy on performing more complex tasks. The original conceptualization of IVA was to aid users in functional tasks (e. g., setting alarms, schedule reminders, checking emails). However, the process itself was quite formal, technical, and mechanized. The results of this study encourage designers to re-imagine the interface and set up IVAs in a manner that individuals feel they are interacting with peers at the workplace. Considering our research found both copresence and faith were related to the comfortable feeling toward Siri as a coworker, the re-design can focus on (1) making the IVA become more familiar to the user's voice possibly through a machine learning technique, (2) expanding the capacity of the device to perform a wide range of basic and intermediate tasks, and (3) ensuring that the conversation algorithm of the IVA responds to user queries in a sensitive and compassionate

The finding about interaction between Siri's gendered voice and participants' gender seems to suggest participants may go through a social identification process leading to a higher level of faith in the same gender voice. Thus, having various types of voice options, not only in gender but also in accents, can facilitate this social identification further. Currently, Siri has options to choose from Australian, British, Indian, Irish, and South African accents. Expanding these options along with different languages may work positively for widening the user base of IVAs and facilitating users' trust toward the AI-system. Based on improved natural language processing and machine learning, if IVAs can mirror our speech patterns and accommodate our communication better, users will feel higher levels of copresence while interacting with them and trust toward them will be strengthened.

# 6. Conclusion

The current study experimented how a common type of intelligent virtual agent, Siri's gendered-voice, its match with participants' gender, and assistance with specific types of task influenced trust and social presence; we further explored how social presence and trust were associated with participants feeling comfortable with Siri as a coworker, supervisor, and friend. Results showed cognitive dimensions of trust (i. e., perceived reliability and understandability, and technical competence) were associated with functional task performance assisted by Siri, but affective dimensions of trust (i.e., faith and personal attachment) were not associated with social task performance. A significant interaction between participant's gender and gender of Siri's voice was identified on the faith dimension of trust. Participants had a higher level of faith on Siri's gendered voice when it matched their own gender, and women participants had lower faith in Siri than men overall. Copresence and affective trust seem to facilitate a more comfortable feeling toward Siri as various relational partners. These research findings contribute to theory building in human-machine communication (HMC) by expanding theoretical components of task characteristics and examining both functional and relational aspects of HMC and their connections.

#### CRediT authorship contribution statement

**Sun Kyong Lee:** Conceptualization, Methodology, Software, Validation, Formal analysis, Writing – original draft, Visualization, Supervision, Resources. **Pavitra Kavya:** Conceptualization, Methodology, Project administration, Data curation. **Sarah C. Lasser:** Writing – review & editing.

#### **Declaration of Competing Interest**

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

#### Appendix I

- Questions for the Functional Task Setting
- Q1. What's the weather like today?
- Q2. Where's the NASDAQ today?
- Q3. What is a 20 percent tip on \$437?
- Q4. What is the square root of 144?
- Q5. How many cups are in a quart?
- Q6. How many days until June 18, 2020?
- Q7. What time is it in Tokyo?
- O8. Define Chillax?
- O9. Where is Starbucks?
- Q10. How much does gas cost right now?
- Q11. What's the nearest Museum?
- Q12. What basketball games are on today?
- Q13. How do I say my name in Mandarin?
- Q14. How do I say how are you doing in German?
- Q15. What airplanes are currently above me?
- Questions for the Social Task Setting
- Q1. Hi Cortana! (Note: Cortana is a virtual assistant created by Microsoft for Windows 10)
  - Q2. What does Siri Mean?
  - Q3. Are you a robot?
  - Q4. Why did Apple make you?
  - Q5. How much do you earn?
  - Q6. How old are you?
  - Q7. Siri, do you sleep?
  - Q8. What's your favorite animal?
  - Q9. What came first, the chicken or the egg?
  - Q10. Who's on first?
  - Q11. What does the fox say?
  - Q12. Will pigs fly?
- Q13. How much wood could a woodchuck chuck if a woodchuck could chuck wood?
- Q14. How many Apple Store geniuses does it take to screw in a light bulb?
  - Q15. What is the best operating system?

#### References

- Biocca, F., Harms, C., 2002. Defining and measuring social presence: Contribution to the networked minds theory and measure. Proceedings of PRESENCE 7–36, 200210.1.1.84.8350.
- Biocca, F., Harms, C., Burgoon, J.K., 2003. Toward a more robust theory and measure of social presence: Review and suggested criteria. Presence: Teleoperators & virtual environments 12 (5), 456–480. https://doi.org/10.1162/105474603322761270.
- Brave, S.B., Nass, C., Hutchinson, K., 2005. Computers that care: Investigating the effects of orientation of emotion exhibited by an embodied computer agent. International Journal of Human-Computer Studies 62, 161–178. https://doi.org/10.1016/j.ijhcs.2004.11.002.
- Cai, Z., Fan, X., Du, J., 2017. Gender and attitudes toward technology use: A metaanalysis. Computers & Education 105, 1–13. https://doi.org/10.1016/j. compedu.2016.11.003.
- Chang, H.H., 2010. Task-technology fit and user acceptance of online auction. International Journal of Human-Computer Studies 68, 69–89. https://doi.org/ 10.1016/i.jihcs.2009.09.010.
- Chou, Ch., Wu, H.Ch., Chen, Ch.H, 2011. Re-visiting college students' attitudes toward the Internet-based on a 6-T model: Gender and grade level difference. Computers & Education 56 (4), 939–947. https://doi.org/10.1016/j.compedu.2010.11.004.
- Daft, R.L., Lengel, R.H., 1986. Organizational information requirements, media richness and structural design. Management Science 32 (5), 554–571. https://doi.org/ 10.1287/mnsc.32.5.554.
- Damen, N., Toh, C., 2019. Designing for trust: Understanding the role of agent gender and location on user perceptions of trust in home automation. Journal of Mechanical Design (6), 141. https://doi.org/10.1115/1.4042223.

- Derntl, B., Finkelmeyer, A., Eickhoff, S., Kellermann, T., Falkenberg, D.I., Schneider, F., Habel, U., 2010. Multidimensional assessment of empathetic abilities: Neural correlates and gender differences. Psychoneuroendocrinology 35 (1), 67–82. https:// doi.org/10.1016/j.psyneuen.2009.10.006.
- Eyssel, F., Hegel, F., 2012. (S)he's got the look: Gender stereotyping of robots. Journal of Applied Social Psychology 42 (9), 2213–2230. https://doi.org/10.1111/j.1559-1816.2012.00937.x.
- Faul, F., Erdfelder, E., Lang, A.-G., Buchner, A, 2007. G\*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. Behavior Research Methods 39, 175–191.
- Fortin, D.R., Dholakia, R.R., 2005. Interactivity and vividness effects on social presence and involvement with a web-based advertisement. Journal of Business Research 58 (3), 387–396. https://doi.org/10.1016/S0148-2963(03)00106-1.
- Gambino, A., Fox, J., Ratan, R.A., 2020. Building a stronger CASA: Extending the computers are social actors paradigm. Human-Machine Communication 1, 71–86. https://doi.org/10.30658/hmc.1.5.
- Gaudiello, I., Zibetti, E., Lefort, S., Chetouani, M., Ivaldi, S., 2016. Trust as indicator of robot functional and social acceptance. An experimental study on user conformation to iCub answers. Computers in Human Behavior 61, 633–655. https://doi.org/ 10.1016/i.chb.2016.03.057.
- Gefen, D., Straub, D., 2004. Consumer trust in B2C e-commerce and the importance of social presence: Experiments in e-products and e-services. Omega 32 (6), 407–424. https://doi.org/10.1016/j.omega.2004.01.006.
- Gong, L., Nass, C., 2007. When a talking-face computer agent is half-human and half-humanoid: Human identity and consistency preference. Human Communication Research 33 (2), 163–193. https://doi.org/10.1111/j.1468-2958.2007.00295.x.
- Groom, V., Nass, C., 2007. Can robots be teammates? Benchmarks in human-robot teams. Interaction Studies 8 (3), 483–500. https://doi.org/10.1075/is.8.3.10gro.
- Guo, J., Tao, D., Yang, C., 2020. The effects of continuous conversation and task complexity on usability of an AI-based conversational agent in smart home environments. In: Long, S., Dhillon, B.S. (Eds.), Man-machine-environment system engineering: Proceedings of the 19th international conference on MMESE, Eds., pp. 695–703. https://doi.org/10.1007/978-981-13-8779-1 79.
- Guzman, A.L., 2017. Making AI safe for humans: A conversation with Siri. In: Gehl, R.W., Bakardjieva, M. (Eds.), Socialbots and their friends, Eds. Digital media and the automation of sociality, New York: Routledge, pp. 69–82.
- Guzman, A.L., Lewis, S.C., 2019. Artificial intelligence and communication: A human-machine communication research agenda. New Media & Society 22 (1), 70–86. https://doi.org/10.1177/1461444819858691.
- Habler, F., Schwind, V., Henze, N., 2019. Effects of smart virtual assistants' gender and language. Paper presented at ACM, Hamburg, Germany. https://doi.org/10.1145/ 3340764.3344441.
- Hancock, P.A., Billings, D.R., Schaefer, K.E., Chen, J.Y., De Visser, E.J., Parasuraman, R, 2011. A meta-analysis of factors affecting trust in human-robot interaction. Human Factors 53 (5), 517–527. https://doi.org/10.1177/0018720811417254.
- Hassanein, K., Head, M., 2006. The impact of infusing social presence in the web interface: An investigation across product types. International Journal of Electronic Commerce 10, 31–55. https://doi.org/10.2753/JEC1086-4415100202.
- Hess, T., Fuller, M., Campbell, D., 2009. Designing interfaces with social presence: Using vividness and extraversion to create social recommendation agents. Journal of the Association for Information Services 10 (12), 889–919. https://doi.org/10.17705/ 1jais.00216.
- Ho, A., Hancock, J., Miner, A.S., 2018. Psychological, relational, and emotional effects of self-disclosure after conversations with a chatbot. Journal of Communication 68 (4), 712–733. https://doi.org/10.1093/joc/jqy026.
- Hoffmann, L., Krämer, N.C., 2013. Investigating the effects of physical and virtual embodiment in task-oriented and conversational contexts. International Journal of Human-Computer Studies 71 (7), 763–774. https://doi.org/10.1016/j. ijhcs.2013.04.007.
- Hu, L., Bentler, P.M., 1999. Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. Structural Equation Modeling: A Multidisciplinary Journal 6 (1), 1–55. https://doi.org/10.1080/ 10705519909540118.
- Jain, M., Kumar, P., Kota, R., Patel, S., 2018. Evaluating and informing the design of chatbots. In: Proceedings of the 2018 Designing Interactive Systems Conference, pp. 895–906. https://doi.org/10.1145/3196709.3196735.
- Karahanna, E., Straub, D.W., 1999. The psychological origins of perceived usefulness and ease-of-use. Information & Management 35 (4), 237–250. https://doi.org/10.1016/ S0378-7206(98)00096-2.
- Kavya, P., Bi, D., Lee, S., 2019. Trusting Siri: Effect of gender and task on human-robot interactions. Paper presented at the annual meeting of the International Communication Association, Washington, D.C.
- Kline, R.B., 2011. Methodology in the Social Sciences. Principles and practice of structural equation modeling, 3rd ed. Guilford Press.
- Lee, E.J., Nass, C., Brave, S., 2000. Can computer-generated speech have gender? an experimental test of gender stereotype. CHI '00 Extended abstracts on human factors in computing systems. Association for Computing Machinery, Hague, The Netherlands, pp. 289–290. https://doi.org/10.1145/633292.633461.
- Lee, K.M., 2004. Presence, explicated. Communication Theory 14 (1), 27–50. https://doi. org/10.1111/j.1468-2885.2004.tb00302.x.
- Lee, K.M., Nass, C., 2004. The multiple source effect and synthesized speech: Doubly-disembodied language as a conceptual framework. Human Communication Research 30 (2), 182–207. https://doi.org/10.1111/j.1468-2958.2004.tb00730.x.
- Lee, J.R., Nass, C.I., 2010. Trust in computers: The computers-are-social-actors (CASA) paradigm and trustworthiness perception in human-computer communication. In:

- Latusek, D., Gerbasi, A. (Eds.), Trust and technology in a ubiquitous modern environment: Theoretical and methodological perspectives. IGI Global, pp. 1–15.
- Lee, K.M., Park, N., 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 (4), 538–563. https://doi.org/10.1111/j.1468-2958.2005.tb00882.x.
- Lee, J.D., See, K.A., 2004. Trust in automation: Designing for appropriate reliance. Human Factors 46 (1), 50–80. https://doi.org/10.1518/hfes.46.1.50\_30392.
- Liao, Q.V., Davis, M., Geyer, W., Muller, M., Shami, N.S., 2016. What can you do? Studying social-agent orientation and agent proactive interactions with an agent for employees. In: Proceedings of the 2016 ACM Conference on Designing Interactive Systems, pp. 264–275. https://doi.org/10.1145/2901790.2901842.
- Lin, C.H., Liu, E.Z.F., Huang, Y.Y. 2012. Exploring parents' perceptions towards educational robots: Gender and socioeconomic differences. British Journal of Educational Technology 43, E31–E34. https://doi.org/10.1111/j.1467-8535.2011.01258.x.
- Madsen, M.A., Gregor, S., 2000. Measuring human-computer trust. In: Proceedings of the 11th Australian conference on information systems. AIS, pp. 53–64.
- McDonnell, M., Baxter, D., 2019. Chatbots and gender stereotyping. Interacting with Computers 31 (2), 116–121. https://doi.org/10.1093/iwc/iwz007.
- Mitchell, W.J., Ho, C-C., Patel, H., Mac-Dorman, K.F., 2011. Does social desirability bias favor humans? Explicit-implicit evaluations of synthesized speech support a new HCI model of impression management. Computers in Human Behavior 27 (1), 402-412. https://doi.org/10.1016/j.chb.2010.09.002.
- Nass, C., Brave, S., 2005. Wired for speech: How voice activates and advances the human-computer-relationship. The MIT Press.
- Nass, C., Moon, Y., Green, C., 1997. Are machines gender-neutral? Gender-stereotypic response to computers with voices. Journal of Applied Social Psychology 27, 864–876. https://doi.org/10.1111/j.1559-1816.1997.tb00275.x.
- Nass, C., Lee, K.M., 2001. Does computer synthesized speech manifest personality?: Experimental tests of recognition, similarity attraction, and consistency-attraction. Journal of Experimental Psychology. Applied 7, 171–181. https://doi.org/10.1037/1076-898X.7.3.171.
- Nomura, T., 2017. Robots and gender. Gender and the Genome 1 (1), 18-25.
- Nowak, K.L., Biocca, F., 2003. The effect of the agency and anthropomorphism on users' sense of telepresence, copresence, and social presence in virtual environments. Presence: Teleoperators & Virtual Environments 12 (5), 481–494. https://doi.org/10.1162/105474603322761289.
- Payne, J., Szymkowiak, A., Robertson, P., Johnson, G., 2013. Gendering the machine: Preferred virtual assistant gender and realism in self-service. In: Aylett, R., Krenn, B., Pelachaud, C., Shimodaira, H. (Eds.), International Workshop on Intelligent Virtual Agents, Eds. Springer, Berlin, Heidelberg, pp. 106–115.
- Reeves, B., Nass, C.I., 1996. The media equation: How people treat computers, television, and new media like real people and places. Cambridge University Press.
- Ruijten, P.A., Midden, C.J., Ham, J., 2015. Lonely and susceptible: The influence of social exclusion and gender on persuasion by an artificial agent. International Journal of Human-Computer Interaction 31 (11), 832–842, 10.108010447318.2015.1067480.
- Rzepka, C., Berger, B., 2018. User interaction with AI-enabled systems: A systematic review of IS research. In: Proceedings of the 39th International Conference on Information Systems. San Francisco, USA.

- Skalski, P., Tamborini, R., 2007. The role of social presence in interactive agent-based persuasion. Media Psychology 10, 385–413. https://doi.org/10.1080/ 15213260701533102.
- Short, J., Williams, E., Christi, B., 1976. The social psychology of telecommunications. Wiley, London.
- Sin, J., Munteanu, C., 2020. An empirically grounded sociotechnical perspective on designing virtual agents for older adults. Human–Computer Interaction. Published online first. https://doi.org/10.1080/07370024.2020.1731690.
- Sundar, S.S., 2008. The main model: A heuristic approach to understanding technology effects on credibility. In: Metzger, M.J., Flanagin, A.J. (Eds.), Digital media, youth, and credibility, Eds. The MIT Press, pp. 73–100.
- Tabachnick, B.G., Fidell, L.S., 2013. Using Multivariate Statistics, 6th edition. Pearson.
- Thórisson, K. R., Bieger, J., Thorarensen, T., Sigurðardóttir, J. S., & Steunebrink, B. R. (2016). Why artificial intelligence needs a task theory and what it might look like. Retrieved from http://arxiv.org/abs/1604.04660.
- Tu, C.H., 2002. The measurement of social presence in an online learning environment. International Journal on E-learning 1 (2), 34–45. https://doi.org/10.17471/2499-4324/421.
- Turkle, S., 2012. Alone together: Why we expect more from technology and less from each other. Basic Books.
- Wang, J., Wang, X., 2012. Structural equation modeling: Applications using Mplus. Wiley, Hoboken, NJ.
- Wirtz, J., Patterson, P.G., Kunz, W.H., Gruber, T., Lu, V.N., Paluch, S., Martins, A., 2018. Brave new world: service robots in the frontline. Journal of Service Management 29 (5), 907–931. https://doi.org/10.1108/JOSM-04-2018-0119.
- Yagoda, R.E., Gillan, D.J., 2012. You want me to trust a ROBOT? The development of a human-robot interaction trust scale. International Journal of Social Robotics 4 (3), 235–248. https://doi.org/10.1007/s12369-012-0144-0.

Sun Kyong Lee (Ph.D., Rutgers University) is an Associate Professor in the Department of Communication at the University of Oklahoma. Her research examines affordances and socio-cultural implications of communication technologies including mobile and social media, and artificial intelligence systems. Examples of her research can be found in Journal of Computer-Mediated Communication, Computers in Human Behavior, and Communication Research.

Pavitra Kavya (Ph.D., University of Oklahoma) is an Assistant Professor in the Department of Communication at California State Polytechnic University, Pomona. Her research interests include leadership communication, training and development, and positive organizational scholarship, with a special emphasis on understanding and helping individuals flourish and thrive at their workplace.

Sarah Lasser (M.A., Valparaiso University) is a doctoral student with research interests in organizational communication, leadership, and gender studies. She is currently a graduate teaching assistant in the Department of Communication at the University of Oklahoma