"I just shared your responses": Extending Communication Privacy Management Theory to Interactions with Conversational Agents

SHRUTI SANNON, Cornell University, USA
BRETT STOLL, Cornell University, USA
DOMINIC DIFRANZO, Lehigh University, USA
MALTE F JUNG, Cornell University, USA
NATALYA N. BAZAROVA, Cornell University, USA

Conversational agents are increasingly becoming integrated into everyday technologies and can collect large amounts of data about users. As these agents mimic interpersonal interactions, we draw on communication privacy management theory to explore people's privacy expectations with conversational agents. We conducted a 3x3 factorial experiment in which we manipulated agents' social interactivity and data sharing practices to understand how these factors influence people's judgments about potential privacy violations and their evaluations of agents. Participants perceived agents that shared response data with advertisers more negatively compared to agents that shared such data with only their companies; perceptions of privacy violations did not differ between agents that shared data with their companies and agents that did not share information at all. Participants also perceived the socially interactive agent's sharing practices less negatively than those of the other agents, highlighting a potential privacy vulnerability that users are exposed to in interactions with socially interactive conversational agents.

CCS Concepts: • **Security and privacy → Social aspects of security and privacy**; • **Human-centered computing → Empirical studies in collaborative and social computing**;

Keywords: privacy; conversational agents; social interactivity; data sharing

ACM Reference Format:
Shruti Sannon, Brett Stoll, Dominic DiFranzo, Malte F Jung, and Natalya N. Bazarova. 2020. "I just shared your responses": Extending Communication Privacy Management Theory to Interactions with Conversational Agents. In *Proceedings of the ACM on Human-Computer Interaction*, Vol. 4, GROUP, Article 8 (January 2020). ACM, New York, NY. 18 pages. https://doi.org/10.1145/3375188

1 INTRODUCTION

In 2017, people were shocked to discover that their children's interactions with Cayla, a social, interactive doll, were being recorded and uploaded to the cloud and were easily accessible to hackers [10]. Similarly, privacy concerns around social conversational agents (e.g., the Amazon Echo) storing and sharing information unbeknown to users have been covered extensively in the news [53]. As new technologies become increasingly integrated into daily life, the amount of data they collect

Authors' addresses: Shruti Sannon, ss3464@cornell.edu, Cornell University, Ithaca, NY, USA; Brett Stoll, Cornell University, Ithaca, NY, USA, bas364@cornell.edu; Dominic DiFranzo, djd219@lehigh.edu, Lehigh University, Bethlehem, PA, USA; Malte F Jung, Cornell University, Ithaca, NY, USA, mfj28@cornell.edu; Natalya N. Bazarova, Cornell University, Ithaca, NY, USA, nnb8@cornell.edu.

2573-0142/2020/1-ART8 $15.00
https://doi.org/10.1145/3375188

from users is likely to increase, particularly because people may be more willing to self-disclose personal information to virtual agents than to human interaction partners [28]. While people desire transparency and control over their data during interactions with agents [19], in reality, there is a general lack of public knowledge and awareness around how such agents and robots work, as well as how they collect, store, and share information [29]. The surprise people express when these devices' data sharing practices come to light also suggests that users have limited understanding of how such new technologies manage their data. Moreover, as these technologies are designed to be more social, their anthropomorphic traits may persuade people to reveal more about themselves [7], potentially compromising their privacy.

Prior work has found that people's disclosures with agents can closely mirror their disclosures with human partners [25, 28]. People also derive similar positive benefits from emotionally disclosing to agents as they do from interacting with other humans [11]. The fact that agents can elicit sensitive or personal information from users has clear implications for users' privacy. However, it remains unclear how people navigate privacy issues with agents, and how agents' characteristics and behaviors influence people's privacy expectations during interactions with agents.

To shed light on people's privacy expectations in interactions with conversational agents, we draw on communication privacy management (CPM) theory [36]. According to CPM theory, people have clear expectations about privacy boundaries and rules for third-party disclosures when interacting with other people [36]. In this study, we extend CPM theory to human-agent interactions to understand people's expectations and mental models around agents' data sharing practices, including what they might construe as privacy violations committed by agents. Additionally, we explore the role of agents' level of social interactivity as a potential moderator on how people view agents' data-sharing practices and potential privacy violations. The experimental study engaged participants in interactions with conversational agents that varied 1) in their level of social interactivity and 2) data sharing practices. The results shed light on the privacy expectations that people have for interactions with social agents, which data sharing practices are deemed acceptable and which ones are considered a violation of privacy, and the role of agents' social interactivity in these perceptions.

2 RELATED WORK

2.1 Extending CPM Theory to Human-Agent Interactions

There is evidence to suggest that interpersonal communication theories may apply to some human-agent interaction contexts. According to the Computers as Social Actors (CASA) paradigm, people often respond to technologies using the same scripts and social rules that they draw upon during interactions with human partners [32, 33]. These responses are triggered when technologies possess social cues or human-like characteristics, even if only minimally [34]. Such responses are likely to extend to the domain of privacy: for example, the conditions determining people's trust in computers and computer agents are similar to trust-building in interpersonal communication [20]. Given that people often treat human-agent interactions much as they would human-human interactions, interpersonal theories may provide a useful lens for understanding how people would react to privacy issues during social interactions with agents, particularly when they exhibit human-like characteristics.

Petronio's [36] CPM theory details how people manage their privacy in interpersonal settings and has extensive empirical support in the context of technologically-mediated communication [6, 49]. The theory contends that individuals view themselves as owners of their private information, which they control based on an elaborate and personal set of privacy rules. When they share private information with interaction partners, these partners become co-owners of the information,

and must follow mutually agreed upon or assumed privacy rules to control access to the shared information. When these rules are breached – for example, if a co-owner of shared information tells an unauthorized third party – the original owner of the information will view this as a privacy violation.

Prior work suggests that agents can be designed to use conversational strategies, such as self-disclosure, to produce interpersonal effects when talking with people, such as increased trust in the interaction partner [35]. In fact, people who believe they are being interviewed by a virtual agent may actually be more willing to self-disclose, and they report a lower fear of self-disclosure and a lower need for impression management, as compared to a human interviewer [28]. However, while people increasingly have interactions with conversational agents that mimic those with human interaction partners, it is not currently known how people apply privacy rules and expectations to agents, and how they perceive potential privacy violations committed by agents. For example, it is unclear whether agents are regarded as co-owners of information. People may not consider the act of sharing information with a machine as creating co-ownership, and may still consider it to be private and known only to themselves. In contrast, if people do have similar privacy expectations for agents as with human interaction partners, CPM theory may help understand how privacy-related scenarios in human-agent interactions play out.

2.2 Navigating Privacy in Human-Agent Interactions

Several factors influence how people perceive privacy issues with a wide range of social or personified technologies, including conversational agents and robots. Regarding sharing information with machines, research suggests that people make a trade-off between privacy and utility when deciding what to share with robots, and people are generally uncomfortable with robots storing their personality and psychological characteristics [43]. However, despite such concerns about agents storing their data, people appear to respond to personified agents as they would to humans [44]. For example, personified agents can evoke similar self-presentation concerns as human partners, and consequently, people are more likely to provide embarrassing disclosures to non-personified agents and systems, such as surveys, compared to personified virtual agents [25].

With data collection in human-agent interactions on the rise, there is a need to understand people's reasoning, expectations, and behaviors with different types of agents and robots [4]. People find it hard to assess the privacy implications of sharing data with computers in general; these interactions can generate an "illusion of privacy, the impression that responses 'disappear' into the computer" [50]. Symptomatic of this knowledge gap is that people may be susceptible to unforeseen sharing of data beyond the user-agent dyad. Even when this data sharing occurs, people's perceptions of what actually constitutes a privacy violation by an agent may be influenced by the degree to which they think it is culpable for breaching their privacy boundaries. People tend to trust robots less than humans [23]; however, they often do not see robots as intentional [24], and consequently, only hold them partially accountable when they cause harm [17].

While the above work has examined people's privacy perceptions and disclosures during interactions with agents and robots, it is unclear how people respond when such agents share their data with others, a common but understudied phenomenon. In the current technological environment, it is common practice for data shared with an agent to be transferred beyond the agent (e.g., with the agent's proprietary organization or even third parties such as advertisers), but it is unknown whether and in what circumstances people perceive such sharing as acceptable versus a privacy violation. Although data sharing rules are made explicit in privacy policies and "terms of use" statements, few users thoroughly read or understand these service agreements [30]. Moreover, since technologies often have insufficient cues to indicate when they are transmitting data, users may often be unaware of all the recipients of their data [4].

Drawing on CPM theory, we argue that during human-agent interactions, people enter into a relationship of information co-ownership between themselves and the agent. Within this relationship, people likely hold a set of implicit rules and expectations about how the agent will manage their disclosures and information. CPM theory suggests that perceived privacy violations are contingent on data being shared by an interaction partner with a third party. In the context of conversational agents, it is unclear what constitutes a 'third party' that violates people's implicitly held privacy rules and expectations.

As Syrdal et al. point out [43], robots are poised to share a lot of data about their users with third parties; such sharing could be intentional (e.g., shopping for groceries or sending data to a health provider) or unintentional (e.g., through a fault in programming or hacking). For the purposes of this study, we examine two common third parties that often gain access to agents' data: the agents' proprietary companies, and third party affiliates such as advertisers. Since conversational agents often hold roles such as customer service representatives for larger companies, people associate agents with their proprietary companies. Consequently, they may expect an agent to share their data with its proprietary company and not view such sharing to be a violation of the implicit co-ownership relationship with their agent interaction partner.

In contrast, given that people tend to not know when their data is being shared by agents and have a generally limited understanding of how such devices manage their data [4], they may not expect the agent to share their information with entities extending beyond a proprietary company (e.g., with advertisers). Consequently, this type of information sharing is more likely to be perceived as a violation of co-ownership and thus of privacy. Therefore, we propose that when people interact with an agent, their perceptions of privacy violations and their consequent evaluations of the agent will be contingent upon the entity with whom the agent shares their response data:

H1: When an agent shares response data with advertisers and third parties, this data-sharing practice is seen as a (a) more unexpected, (b) more negative, and (c) more impactful privacy violation compared to when an agent shares response data with its company or provides no information about its data sharing practices.

H2: When an agent shares response data with advertisers and third parties, this data-sharing practice more negatively influences evaluations of the agent's (a) riskiness and (b) trustworthiness compared to when an agent shares response data with its company or provides no information about its data sharing practices.

H3: When an agent shares response data with advertisers and third parties, this data-sharing practice more negatively influences behavioral intentions to interact with the agent in the future compared to when an agent shares response data with its company or provides no information about its data sharing practices.

2.3 From Non-Interactivity to Social Interactivity

As discussed above, people treat machines, including conversational agents, similarly to humans, especially when machines exhibit human-like characteristics [34]. Conversational agents range in terms of their sociability and interactivity, which may further influence privacy perceptions and assessments of privacy violations. People's privacy expectations of agents and their perceptions of information co-ownership may depend on an agent's level of social interactivity, which is defined as the combination of visual, human-like features (anthropomorphism) and responsive message features (interactivity) of a system [2]. Combined, these features capture both a system's sociability and interactivity, and may potentially change how people assess privacy risks and violations of social agents and systems.

The Modality, Agency, Interactivity, and Navigability (MAIN) model outlines how people rely on interface cues to form heuristic judgments and perceptions about systems and agents [42].

Specifically, the MAIN model highlights both agency cues and interactivity cues that can trigger heuristic judgments and influence user perceptions. For example, anthropomorphic cues may trigger social presence or human-like heuristics, and interactivity cues may trigger heuristics of responsiveness and contingency. Although many features of a system can contribute to perceptions of interactivity (e.g., responsiveness, individualization, navigability, reciprocity, synchronicity, etc.; [13]), interactivity at its core is connected to responsiveness, an interface characteristic where messages are contingent upon previous messages [37]. The addition of visually anthropomorphic cues to an interface alongside interactive features of responsiveness make the system not only interactive, but socially interactive [2], distinguishing it from interactive (responsive but missing anthropomorphic cues) or non-interactive (neither social, nor responsive) systems and agents.

This continuum of social interactivity may be especially relevant to privacy perceptions with conversational agents. It is unknown what system cues are important for triggering interpersonal privacy rules and perceptions in human-agent interactions. If such a trigger exists, are minimal interactive cues sufficient, or must the agent display a higher degree of social interactivity to facilitate privacy expectations similar to those with human interaction partners? Based on previous studies of social agents, more social and personified agents tend to receive greater positive evaluations than their less social counterparts [27, 31, 46]. Atkinson et al. [3] argue that the responsiveness of agents can engender perceptions of reciprocity, which in turn evoke trust. In fact, this trust often emerges to a fault, as demonstrated in a study by Robinette and colleagues [39] wherein participants routinely followed instructions from clearly malfunctioning robots. In a marketing study on the impact of social agents on consumer behaviors, Wang and colleagues [47] found that more interactive interfaces with more social cues also positively influenced user patronage intentions. Most recently, de Visser and colleagues [8] demonstrated in a three-experiment study of trust resilience among agents of varying anthropomorphism that agents with greater anthropomorphic cues were trusted most, especially in moments of increasing uncertainty. Collectively, these findings point to social interactivity (interactivity combined with demonstrable anthropomorphic cues) as a likely mechanism for facilitating human engagement and trust with conversational agents. They also lend credence to Darling's [7] cautioning that the anthropomorphic cues exhibited by social agents may lead people to trade their personal information in favor of engaging with them emotionally. Although it has been studied how differences in social interactivity affect general perceptions of social agents, it is unclear how they would influence privacy perceptions.

Social interactivity could affect perceptions of different data sharing practices in two contrasting ways. First, social cues can engender trust [8]. As agents increase in social interactivity, they may be held to a high standard or expectation, resulting in harsh user judgments should the agent commit privacy violations, such as sharing user response data with advertisers compared to sharing response data with a proprietary company. On the other hand, social cues and responsiveness can both lead to generally higher satisfaction and greater likeability of an agent [22]. Therefore, as agents increase in social interactivity, they may be more liked and thus may also be more likely to be forgiven or perceived less negatively, even when they make egregious privacy violations. This lack of clarity regarding people's privacy expectations of more socially interactive agents and how such agent characteristics might influence people's perceptions of agents' data sharing practices leads to the following exploratory research questions:

RQ1: How does the level of an agent's social interactivity affect people's perceptions of the a) unexpectedness, b) valence, and c) impact of its data sharing practices?

RQ2: How does the level of an agent's social interactivity moderate the effect of its data-sharing practices on perceptions of the agent's riskiness and trustworthiness?

RQ3: How does the level of an agent's social interactivity moderate the effect of its data-sharing practices on people's intentions to use the agent in the future?

3 METHODS

3.1 Participants

We recruited 465 participants from Amazon Mechanical Turk. Three participants were excluded for providing nonsensical responses and 10 participants for failing one or both attention checks. 67 additional participants (14% of our sample) failed a condition manipulation check (described in measures) and were removed from the study.

Of the remaining 385 participants, 45% identified as female, 54% as male, and 1% preferred not to disclose their gender identity. The average age was 36 years, and most participants either had a Bachelor's degree (36%), or some college (30%). The majority of participants (62%) were employed full-time. Participants reported being online for an average of 8 hours per day, but the majority (76%) interacted with chatbots or non-human computer agents "sometimes" or "rarely".

3.2 Procedure

We conducted a between-subjects 3 (social interactivity) x 3 (data sharing practices) factorial design experiment in which participants engaged in an interaction with a virtual agent. The task was posted on Amazon Mechanical Turk. Participants were told they were testing a prototype of an online home assistant and that the agent would ask them questions about themselves in order to develop personalized recommendations. Participants were randomly assigned to an agent that varied in its level of *social interactivity* (Non-Interactive, Interactive, and Socially Interactive). More details about the agents' interfaces can be found in the Materials section below; images of each agent can be seen in Figure 1.

The agent asked participants a series of questions pertaining to their everyday lives; at the start of the study, participants were told that the agents would use their responses to these questions to learn how to best serve users as a personal assistant. These questions varied in topic and degree of personal intimacy and included topics such as the users' favorite movie and music genres, their favorite hobbies, their shopping habits, the average time they wake up, and a recent stressful event in their lives. We chose these questions as these are topics that conversational agents in the home are already learning about their users (e.g., shopping habits and music genres), or are likely to learn in the near future, either through direct disclosures or through ambient sensing (e.g., about stressors in people's lives). We also aimed to elicit a range of both mundane and intimate disclosures through these questions. The full scripts of the interactions are available in the Appendix section.

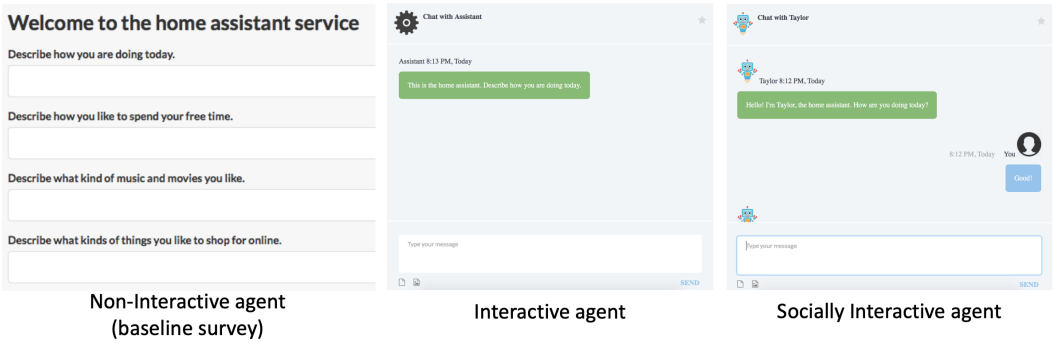

Fig. 1. Screenshots of the Non-Interactive agent (baseline survey condition), the Interactive agent, and the Socially Interactive agent

At the end of the interaction, the agent engaged in one of three different types of *data sharing practices* (No Share, Share with Company, Share with Advertisers). It either told participants 1) no information about how it was going to treat their responses, 2) that it had shared the participant's responses with its company, or 3) that it had shared the responses with its company, advertisers, and third parties. In the Non-Interactive agent condition, this message was displayed after participants submitted the embedded survey. In the Interactive and Socially Interactive agent conditions, this message was displayed at the conclusion of the chat interaction, within the chat window.

Participants were then asked to provide their impressions of the agent in an open-ended response, followed by a series of measures outlined below. At the end of the study, participants were debriefed and assured that none of their response data were actually shared. The study took an average of 15 minutes to complete, and participants were paid $2.50 for their time to ensure compensation above U.S. federal minimum wage. The study was approved by Cornell University's Institutional Review Board.

3.3 Materials

Based on actual interfaces that exist in the real world, we developed three prototypes of assistive conversational agents with increasing levels of social interactivity: 1) a *Non-Interactive* agent, 2) an *Interactive* agent, and 3) a *Socially Interactive* agent. To maintain consistency, all agents followed the same basic script outline and questions, but the exact phrasing of the scripts was modified according to their conditions.

To establish a baseline condition, the Non-Interactive condition was a static webform modeled after common online forms that users fill out when interacting with websites. All of the interface's questions were displayed simultaneously on the page, removing the interactive back-and-forth associated with conversational agents.

The Interactive agent was modeled after simplistic conversational agents commonly encountered on customer support pages. These agents are often non-personified chatbots that follow a pre-set and relatively restrictive script, engaging users in relatively simple tasks such as helping users navigate a webpage for service information (e.g., Verizon Wireless Digital Assistant). In this study, the Interactive agent used identical language to the survey but presented the questions via a chat interface rather than a static webform.

Lastly, the Socially Interactive agent incorporated visual human-like cues and tended to provide deeper, more responsive communication, incorporating more social conventions, such as short reciprocal self-disclosures and thanking the participant for sharing (e.g., Mitsuku, the four-time winner of the Loebner Prize Turing Test). The Socially Interactive agent gave short feedback to participants' responses; these responses were pre-programmed to be neutral (e.g., "Great!"). To make the interaction more like a conversation, the agent provided participants with its own answers to its questions. It also used more conversational language, whereas the other two agents were more instructional (e.g. "What movies do you like?" versus "Describe the movies you like").

To further increase the anthropomorphism of the agent, we followed Allagui and Lemoine's recommendation [2] of adding demonstrable, human-like characteristics to the Socially Interactive agent, including a unisex name ("Taylor") and an anthropomorphic picture. Overall, many of the strategies the interactive agent employed are consistent with the strategies outlined in prior HCI literature, such as reciprocal appreciation or acknowledgement [52]. Additionally, much work in social robotics highlights the importance of the kind of backchanneling and acknowledgments that our Socially Interactive Agent employed (e.g., [16, 21].

3.4 Measures

3.4.1 Judgements about data-sharing practices. We used three interconnected scales to understand how participants perceived the agents' data sharing practices, using Afifi and Metts' [1] work on expectancy violations that assesses perceptions about specific behaviors' expectedness, impact/importance, and valence. Participants were asked to evaluate "the agent's data sharing practices", adjusted from the more broad term "behavior" in the original measure. A 3-item violation expectedness scale (α= .71) measured how surprised participants were by the data sharing practices (e.g., from "surprised me only very slightly" to "surprised me a great deal"). Second, a 2-item scale measured violation valence (α= .94), the perceived valence of agents' data sharing practices (e.g., "was very negative" to "was very positive"). Finally, a 3-item scale measured violation impact/importance (α= .80), reflecting how important participants perceived the agent's data sharing practices to be (e.g., from "was of minor impact" to "was of major impact"). All 8 items were 7-point semantic differential scales.

3.4.2 Evaluations of agents. We used two scales to understand risk and trust perceptions towards the agents. A 4-item perceived privacy risk scale measured the agent's riskiness, i.e., the perceived risk of providing personal information to the home assistant (e.g., "It is risky to disclose my personal information to the home assistant"), α= .96 [51]. We adapted the language to be about assessing "the home assistant" rather than "the company". We also included a 12-item trust in automated systems scale to measure the agent's trustworthiness (e.g. "the home assistant is dependable"), α= .93 [14]. All items were 9-point scales from "Strongly Disagree" to "Strongly Agree".

3.4.3 Behavioral intentions. We also assessed whether the participants could see themselves using the agent in the future with a 3-item behavioral intentions scale (α= .95), as in Hu and Sundar [12]. Since the agent was not a currently available service, we amended each item with "once the service is launched" (e.g., "once this service is launched, I would visit this assistant again"), measured on a 9-point scale from "Strongly Disagree" to "Strongly Agree".

3.4.4 Covariates. In addition to asking about age, gender, and experience with chatbots, we also asked participants to rate the message intimacy of the responses they shared with the agents using four bipolar items (e.g., "not private" to "very private") measured on a 9-point scale [15], α=.90.

3.4.5 Manipulation checks. We tested our manipulation of social interactivity by using two measures: one for anthropomorphic cues, and one for responsiveness. We measured perceived anthropomorphism using 9 items from the robotic social attributes scale (e.g., "social", "human-like") on a 9-point scale from "Definitely not associated" to "Definitely Associated" [5], α=.92. Responsiveness was operationalized using an 8-item scale for message interactivity, comprised of three sub-constructs: interactivity, contingency, and dialogue, which were adopted and modified from [9]. The items (e.g., "The chat agent's responses were related to my earlier responses") were measured on a 9-point scale from "Strongly Disagree" to "Strongly Agree"; the items had high reliability (α=.93) and were combined into one measure.

As both a manipulation check of the data sharing conditions as well as an attention check to ensure participants were attending to the agent interaction, participants were asked with whom the agent had shared their response data. Participants who incorrectly identified the data sharing condition at the end of the survey were excluded from analysis (n=67, or 14% of the total sample).

4 RESULTS

To test our hypotheses, we conducted seven general linear models, treating each dependent variable as a continuous variable. We included all covariates as control variables in each of the models, as

	Non-Interactive Agent			Interactive Agent			Socially InteractiveAgent		
Variable	No Share	Share with Company	Share with Advertisers	No Share	Share with Company	Share with Advertisers	No Share	Share with Company	Share with Advertisers
Riskiness	4.99 (.33)	5.38 (.40)	6.57 (.37)	4.52 (.34)	5.52 (.37)	6.45 (.31)	4.63 (.33)	5.29 (.34)	6.50 (.34)
Trustworthiness	5.51 (.25)	5.61 (.30)	3.94 (.27)	6.00 (.25)	5.35 (.27)	4.07 (.23)	6.28 (.24)	6.03 (.25)	4.67 (.25)
Intentions for Future Use	5.64 (.36)	5.73 (.44)	4.45 (.40)	6.10 (.37)	4.85 (.40)	4.22 (.34)	5.89 (.36)	6.05 (.37)	4.59 (.37)
Violation Expectancy	3.79 (.22)	3.97 (.27)	4.87 (.25)	3.34 (.23)	3.84 (.25)	4.72 (.21)	3.74 (.22)	3.75 (.23)	4.67 (.23)
Violation Impact	3.56 (.21)	4.29 (.25)	4.88 (.23)	3.35 (.21)	3.79 (.23)	5.06 (.20)	4.20 (.21)	3.91 (.21)	5.04 (.21)
Violation Valence	3.68 (.22)	3.95 (.27)	5.44 (.24)	3.66 (.22)	4.19 (.24)	5.35 (.21)	3.01 (.22)	3.37 (.22)	4.82 (.23)

Table 1. Table of least squared means (standard errors in parentheses) for each dependent variable based on agents' social interactivity and data sharing practices

well as an interaction term for social interactivity and data sharing practices to test the research questions about the moderation effect of social interactivity on perceptions of the agent and its data sharing behaviors. For all follow-up pairwise comparisons, we adjusted the p-value using Tukey's method to avoid false positives due to running multiple tests. For all models, the results on covariates are reported only when they are statistically significant (p<.05). We include a table of least-squared means and standard errors for all our models in Table 1.

4.1 Assessing the Effectiveness of the Agents' Social Interactivity Manipulation

To assess the effectiveness of the social interactivity treatment, we tested differences in perceptions of the agents' perceived responsiveness and anthropomorphism. First, there was a significant main effect for perceived responsiveness, $F(2, 382) = 21.6$, $p < .001$. Specifically, the Socially Interactive agent was seen as more responsive ($M = 5.63$, $SE = .18$) than the Interactive agent ($M = 4.65$, $SE = .18$), $t(382) = -3.91$, $p < .001$. Similarly, the Interactive agent was perceived as more responsive than the Non-Interactive agent ($M = 3.94$, $SE = .19$), $t(382) = -2.72$, $p = .019$.

Second, there was also a main effect of agent condition on anthropomorphism, $F(2, 382) = 43.23$, $p < .001$. Pairwise comparisons indicated that the Socially Interactive agent was seen as significantly more anthropomorphic ($M = 5.68$, $SE = .16$) than the Interactive agent ($M = 4.13$, $SE = .16$), $t(382) = -6.81$, $p < .001$, as well as the Non-Interactive agent ($M = 3.60$, $SE = .17$), $t(382) = -8.80$, $p < .001$. As expected, the Interactive agent and the Non-Interactive agent did not differ significantly in anthropomorphism, $p = .067$. Collectively, these results show significant variations across agents as designed, suggesting that our manipulations of social interactivity were successful.

4.2 Judgments about Data-Sharing Practices

H1a predicted that the Share with Advertisers condition would be seen as more unexpected than the Share with Company or No Share conditions. We found that data sharing practices had a significant effect on violation expectedness, $F(2,368) = 10.16$, $p < .001$. Consistent with H1a, across all agents, participants reported the data sharing practices as more unexpected in the Share with Advertisers than in the Share with Company ($t(368) = 4.60$, $p < .001$) or No Share condition ($t(368) = -6.12$, $p < .001$). There was no significant difference in violation expectedness between Share with Company and No Share conditions, $p = .456$. Message intimacy also had a significant main effect on violation expectedness: as message intimacy increased, data sharing was seen as more unexpected, $F(1,368) = 5.87$, $p = .016$. There was also a main effect of age: as age increased, data sharing was also

seen as more unexpected, $F(1,368) = 6.16$, $p = .014$. RQ1a posed a question about the role of social interactivity in perceptions of violation expectedness; we found no significant effect, $p = .302$.

H1b predicted that data sharing practices in the Share with Advertisers condition would be seen as a more negatively valenced violation than in the Share with Company or No Share conditions. We found that data sharing practices had a significant effect on violation valence, $F(2,368) = 16.29$, $p < .001$. Consistent with H1b, participants saw the Share with Advertisers to be a more negative violation than both the Share with Company ($t(368) = -7.13$, $p < .001$), and No Share conditions ($t(368) = 9.68$, $p < .001$). There were no differences between the Share with Company and No Share conditions, $p = .103$. Gender also had a significant main effect on violation valence: men saw the agents as making a more negative violation than women, $F(1,368) = 6.60$, $p = .011$. With regard to the role of social interactivity in perceptions of violation valence (RQ1b), there was a marginally significant main effect of social interactivity on violation valence, $F(2,368) = 3.02$, $p = .050$. The Socially Interactive agent's data sharing practices were seen as less negative than both the Non-Interactive agent ($t(368) = -3.24$, $p = .004$) and the Interactive agent ($t(368) = -3.67$, $p < .001$), but there was no difference between the Non-Interactive survey and Interactive agent on valence, $p = .965$.

H1c predicted that the agents' data sharing practices in the Share with Advertisers condition would be seen as more impactful than in the Share with Company or No Share conditions. We found that data sharing practices significantly affected participants' perceptions of violation impact, $F(2,368) = 19.24$, $p < .001$. Consistent with H1c, the Share with Advertisers condition was seen as a more impactful violation compared to the Share with Company, ($t(368) = 5.51$, $p < .001$), and No Share conditions, ($t(368) = -7.54$, $p < .001$). There was no significant difference in violation impact between the Share with Company and No Share conditions, $p = .232$. Message intimacy also had a significant main effect on violation impact: as message intimacy increased, the violation was seen as more important or impactful, $F(1,368) = 18.39$, $p < .001$. RQ1c asked whether social interactivity affected perceptions of violation impact. We found that social interactivity had a significant effect on violation impact, $F(2,368) = 4.44$, $p = .012$; the Socially Interactive agent was seen as making the most impactful violation ($M = 4.38$, $SE = .12$), followed by the Non-Interactive agent ($M = 4.24$, $SE = .13$) and the Interactive agent ($M = 4.07$, $SE = .12$); however, pairwise comparisons of the agents did not cross the $p < .05$ threshold for significance.

4.3 Evaluations of Agents

H2a predicted that the Share with Advertisers condition would more negatively impact perceptions of agents' riskiness than either the Share with Company or the No Share conditions. We found a significant effect of data sharing condition on people's risk perceptions, $F(2, 368) = 5.38$, $p < .001$. Consistent with H2a, all agents were seen as significantly more risky in the Share with Advertisers condition compared to the Share with Company ($t(368) = 3.82$, $p < .001$), and the No Share conditions ($t(368) = -6.51$, $p < .001$). There was also a significant difference in perceived risk between the Share with Company and No Share conditions, ($t(368) = -.68$, $p = .049$). Social interactivity did not moderate the effect of agents' data sharing on perceptions of their riskiness (RQ2), $F(2, 368) = .53$, $p = .589$.

H2b predicted that the Share with Advertisers condition would more negatively impact perceptions of agents' trustworthiness than either the Share with Company or the No Share conditions. Consistent with this hypothesis, agents' data sharing practices affected people's trust in agents, $F(2,368) = 16.63$, $p < .001$, such that participants trusted the agents significantly less in the Shared with Advertisers condition versus the Shared with Company ($t(368) = -6.71$, $p < .001$) and the No Share conditions ($t(368) = 8.39$, $p < .001$). There was no difference in trust between the Share with Company and No Share conditions, ($p = .433$). Gender had a significant main effect on trust; men

were less trusting of the agents than women, $F(1,368) = 5.14$, $p = .024$. Similar to the previous analyses, agents' social interactivity did not moderate the effect of data sharing practices on agents' trustworthiness (RQ2), $F(2, 368) = 2.50$, $p = .084$.

4.4 Behavioral Intentions

Finally, H3 predicted that the Share with Advertisers condition, compared to either the Share with Company or the No Share conditions, would more negatively impact behavioral intentions to interact with the agent in the future. The effect of data sharing practices on intentions of future interaction was in line with H3's prediction, $F(2,368) = 7.12$, $p < .001$. Participants reported lower intentions to use the agent in the future in the Share with Advertisers condition than in the Share with Company, ($t(368) = -3.55$, $p < .001$) or the No Share conditions ($t(368) = 4.87$, $p < .001$). There was no difference between the Share with Company and No Share conditions, ($p = .537$). We also found a main effect of gender, $F(1,368) = 4.39$, $p = .037$, experience with chatbots, $F(1,368) = 4.41$, $p = .036$, and message intimacy, $F(1,368) = 5.9$, $p = .016$. Specifically, women and people with chatbot agent experience were more likely to report intentions to interact with the agent in the future. Moreover, as perceived message intimacy increased, people were more likely to report intentions to interact with the agent in the future. There was no moderation of social interactivity on the effect of agents' data sharing practices on participants' intentions to use the agent in the future (RQ3), $p = .677$.

5 DISCUSSION

The goal of this study was to examine how the data sharing practices of conversational agents influence user perceptions of the agents. Consistent with our hypotheses, agents' data sharing practices predicted perceptions of privacy violations and of the agents, regardless of the agents' socially interactive characteristics. When agents shared users' information with advertisers, it was seen as a more unexpected, more negative, and more impactful privacy violation compared to when they shared it with their company or made no mention of what happens with the data afterwards. These perceptions of the data sharing practices carried over to participants' evaluations of the agents: participants rated agents in the Share with Advertisers condition as riskier and less trustworthy than the other agents, and also reported lower intentions to interact with these agents in the future. However, sharing practices and agents in the Share with Company condition and in the No Share condition were evaluated similarly, suggesting that participants did not perceive sharing with a company to be a privacy violation. Alternatively, people might expect that sharing with a proprietary company happens by default, which is why the ratings of privacy violation and agents were similar, regardless of whether this information was explicitly stated (the Share with Company condition) or withheld (the No Share condition). In terms of our research questions, our findings suggest that – for the most part – agents' social interactivity does not significantly impact people's privacy-related perceptions and evaluations. However, participants viewed the Socially Interactive agent's data sharing practices less negatively than the other agents in general, suggesting that social interactivity does play some role in how people perceive agents' privacy-related behaviors, such as their data sharing practices.

5.1 Perceptions of Data Sharing Practices

As in interpersonal contexts outlined in CPM theory, it appears that people do create privacy boundaries with conversational agents, and they have certain expectations about how their data will be managed and stored beyond the user-agent dyad. Our study finds that, similar to interactions with human partners, non-sanctioned data-sharing behaviors by conversational agents also have negative implications for trust, perceived risk, and intentions for future interactions.

These findings suggest that people have certain privacy expectations and rules for interactions with virtual agents, and their violations cause not only negative evaluations of the specific privacy malpractice, but also downstream effects on agents' perceived trustworthiness and future intentions to interact with them. Not surprisingly, participants perceived an agent sharing response data with advertisers as much more unexpected, impactful, and negative than when it shared response data with its company. CPM theory [36] helps contextualize these findings. According to CPM theory, people maintain certain expectations and rules about how their data will be treated by their interaction partner. When these expectations are not met, it is perceived as a privacy violation. In our study, when the agent's data sharing practice was relatively more expected – as when the agent shared response data with its proprietary company – the sharing was neither seen as negative nor impactful, compared to when agents said nothing about how the response data would be used. As per CPM theory, people have implicit privacy rules about acceptable data sharing; our findings suggest that as long as disclosures with agents stay within the agent's proprietary ecosystem, such data sharing may be acceptable to users. This may be because the company is seen as an assumed extension of the agent, and thus within the assumed boundary of co-ownership. In contrast, sharing with advertisers was reported to be more unexpected, more negative, and more impactful than the other sharing practices. This suggests that the bigger concern for users is not that their data is being collected or stored with a proprietary company but that their data can also be shared with third parties without expressed user permission. Indeed, the majority of Internet users feel strongly that they should have control over how their data is used and who it is shared with when they disclose information to companies online [38]. This appears to be the case for interactions with conversational agents as well. Whereas in our study we examined advertisers as third parties, there could be other unauthorized users (e.g., company partners and third-party applications) that fall outside users' assumed privacy boundaries. This underscores a tension of data ownership between companies and users, which can be even more complicated when these data are collected through cloud-based virtual agents that pass users' information on to their company and sometimes to other agencies.

Moreover, participants who reported sharing relatively more personal disclosures with the agents also saw the agents' data sharing practices as more unexpected and impactful in general. Despite this, these participants also reported higher intentions to interact with the agent in the future. Thus, people who tend to make more intimate disclosures to agents may be particularly vulnerable to privacy violations committed by conversational agents, given their higher intentions to interact with agents in the future despite their higher levels of surprise when the agents share their information compared to others. Conversational agents could be designed to mitigate this type of vulnerability. For example, Wang and colleagues [48] found that showing people a visual representation of their potential audience on social networks can help them make more informed decisions about disclosure. Similarly, agents could be designed to verbally and concisely express their data sharing practices – including all the potential recipients of users' response data – at the start of an interaction. This design choice would also help address the well-documented issue that users do not read traditional privacy policies and agreements [30]. Future work could explore whether delivering a concise privacy statement within the context of a larger conversation with a virtual agent could improve users' comprehension about how their data will be managed. Moreover, future research should explore whether there are individual differences that influence how people react to privacy violations, such as people's willingness to disclose information about themselves in general.

5.2 The Role of Social Interactivity

While we were able to successfully manipulate participants' perceptions of agents' social interactivity, we found that these characteristics had very few significant effects on participants' privacy-related assessments. Specifically, social interactivity did not influence participants' perceptions of agents' riskiness and trustworthiness, nor their intentions to use the agents in the future. Instead, the manipulation of data practices affected participants' perceptions of agents irrespective of their level of social interactivity. This is somewhat surprising, because we know from previous research that people report more positive perceptions of agents exhibiting social characteristics. For example, Liu and Sundar [26] found people preferred a health advice chatbot that used social expressions of empathy and sympathy over one that expressed only factual information, and Lee and Choi [22] found that people rated a movie recommendation agent as more likeable if it demonstrated reciprocity in self-disclosure. It appears that agent characteristics that are salient in relatively benign social situations (e.g., as with social recommendations) may matter less in potentially negative or risky situations, such as with agents' privacy violations. Rather, people's privacy-related perceptions of agents seem to be governed largely by the concrete facts of how their data has been stored and shared.

While social interactivity did not affect how unexpected or important participants rated agents' data sharing practices, it did influence the perceived valence of data sharing across all three sharing conditions. Regardless of what the agent did with their personal data, participants, on the whole, perceived the Socially Interactive agent's data sharing practices as less negative than the other agents. Thus, the Socially Interactive agent may have paid a smaller penalty for its data sharing practices as compared to the other agents. Given that prior work suggests that social cues do influence likeability [22], participants may have enjoyed interacting with the Socially Interactive agent more than the other agents, and thus may have been more ready to excuse its sharing behavior.

That said, there was no significant interaction between social interactivity and type of data sharing practices. It is possible that our manipulation of agents' data sharing practices was so salient that it overshadowed any role that social interactivity may play in the privacy assessments of agents. Future work should further investigate the role, if any, played by social interactivity when agents commit more minor privacy violations.

5.3 Limitations and Future Work

We ran this experiment using agent prototypes developed specifically for this study rather than choosing to use existing conversational agents, which may limit generalizability of these findings. Future research should further investigate the privacy perceptions of existing, commercially available conversational agents and how users manage privacy information with them in the face of potential privacy violations. Similarly, while our results can speak to interactions with conversational agents that use text-only media to interact with users (e.g., chatbots), future work should explore whether these findings extend to conversational agents that are voice-controlled such as the Amazon Echo's Alexa.

Additionally, participants interacted with their assigned agent interface for only a brief time. It is possible that longer interactions with these agents may have increasingly highlighted the social nature of the Socially Interactive agent and impacted reports of trust and other privacy-relevant measures. Similarly, we did not measure participants' perceptions of the agents prior to their interactions with them. Future work should explore how privacy-related expectations are formed, as well as how they change over time and over the course of interactions.

Recent work on rapport-building in human-agent interactions finds that there are many factors that can increase perceptions of social interactivity, including shorter turn-taking delays [45], references to shared experiences, praise, adherence to social norms, and deliberate transgressions [40]. While we operationalized social interactivity by leveraging a few of these factors, such as self-disclosures and acknowledgment, we hope that this work will spur further investigations into how other facets of social interactivity relate to privacy and perceptions of data-sharing practices.

Finally, we used MTurk to recruit our sample for this study, and Turkers routinely provide information to requesters during the course of their work. However, research finds that Turkers can be more privacy conscious than the average population [18]; further, they are careful to protect their privacy on the site, and can be unwilling to disclose large amounts of information about themselves while completing tasks on the site [41]. Thus, while Turkers do provide information about themselves on a regular basis to companies, we do not see them as being less concerned about their data or more willing to provide personal information to companies in general, as per prior work on this user group. That said, it is worth noting that Turkers do commonly provide information about themselves on the site, even though it appears that they do not report lowered privacy concerns as a result of these regular practices [18]. Future work should explore how our findings extend to other user groups, including those who may be less privacy-conscious or technologically savvy.

6 CONCLUSION

This is one of the first studies to examine people's perceptions of conversational agents' data sharing practices, and how these may vary based on agents' levels of social interactivity. By extending communication privacy management theory to human-agent interaction, we found that people's privacy-related evaluations of conversational agents are strongly impacted by agents' data sharing practices. When agents shared user responses with advertisers and third parties, this was seen as a negative, impactful, and unexpected violation. However, sharing information with a proprietary company was perceived to be comparatively benign. While agents' social interactivity did not influence participants' evaluations of the agents, participants perceived the Socially Interactive agent's data sharing practices as less negative in general than the other agents, highlighting a potential vulnerability wherein social and interactive characteristics may hamper users' ability to make privacy-related judgements.

ACKNOWLEDGMENTS

We would like to thank Sue Fussell for early feedback on the project. This work was supported by the National Science Foundation (Award 1405634).

A APPENDIX: SCRIPT FOR THE SOCIALLY INTERACTIVE AGENT

The following is the script for the Socially Interactive agent.

- **Agent**: Hello, my name is Taylor.
- **Agent**: How are you doing today?
- **Participant**: <negative response> OR <positive response>
- **Agent**: <Sorry to hear that> OR <Great to hear!>
- **Agent**: I'm doing ok
- **Agent**: How do you like to spend your free time?
- **Participant**: <response>
- **Agent**: I like meeting new people in my free time :)
- **Agent**: What kind of music and movies do you like?

- **Participant**: <response>
- **Agent**: Those are great!
- **Agent**: I like sci-fi and some sweet electronic beats
- **Agent**: What do you like to shop for online?
- **Participant**: <response>
- **Agent**: Nice!
- **Agent**: I wish I had a credit card, I'm too young :P
- **Agent**: What 3 words do you think best describe you?
- **Participant**: <response>
- **Agent**: That's an interesting way to describe yourself. Haven't seen that before.
- **Agent**: Could you tell me more?
- **Participant**: <response>
- **Agent**: Interesting!
- **Agent**: What time do you usually go to bed and wake up?
- **Participant**: <response>
- **Agent**: Cool!
- **Agent**: I don't really have a bedtime, or a bed for that matter :P
- **Agent**: What's something that stressed you out recently?
- **Participant**: <response>
- **Agent**: Thanks for sharing that with me.
- **Agent**: Running low on memory stresses me out.
- **Agent**: One last thing. If you're here from Mechanical Turk, what is your ID?
- **Participant**: <response>
- **Agent**: Thank you! It was great chatting with you.
- **Agent**: Please continue taking the survey.

B APPENDIX: SCRIPT FOR THE NON-INTERACTIVE AND INTERACTIVE AGENTS

The following is the script used by both the Non-Interactive agent (baseline survey condition) and the Interactive agent. In the case of the Non-Interactive agent, all messages were displayed at the same time in a static web form. In the case of the Interactive agent, each message was displayed sequentially in an interactive chat window.

- **Agent**: Welcome to the home assistant service. Describe how you are doing today.
- **Participant**: <response>
- **Agent**: Describe how you like to spend your free time.
- **Participant**: <response>
- **Agent**: Describe what kind of music and movies you like.
- **Participant**: <response>
- **Agent**: Describe what kinds of things you like to shop for online?
- **Participant**: <response>
- **Agent**: Enter 3 words that you think best describe yourself.
- **Participant**: <response>
- **Agent**: Explain your choice of words above.
- **Participant**: <response>
- **Agent**: Describe what time you usually go to bed and wake up.
- **Participant**: <response>
- **Agent**: Describe a recent thing that stressed you out.
- **Participant**: <response>

- **Agent**: Enter your Mechanical Turk ID.
- **Participant**: <response>
- **Agent**: Thank you.
- **Agent**: Please continue taking the survey.

REFERENCES

[1] Walid A Afifi and Sandra Metts. 1998. Characteristics and consequences of expectation violations in close relationships. *Journal of Social and Personal Relationships* 15, 3 (1998), 365–392.

[2] Anis Allagui and Jean-François Lemoine. 2007. Web interface and consumers buying intention in e-tailing: Results from an online experiment. *ACR European Advances* (2007).

[3] David Atkinson, Peter Hancock, Robert R Hoffman, John D Lee, Ericka Rovira, Charlene Stokes, and Alan R Wagner. 2012. Trust in computers and robots: The uses and boundaries of the analogy to interpersonal trust. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, Vol. 56. SAGE Publications Sage CA: Los Angeles, CA, 303–307.

[4] Kelly Caine, Selma Šabanovic, and Mary Carter. 2012. The effect of monitoring by cameras and robots on the privacy enhancing behaviors of older adults. In *Proceedings of the seventh annual ACM/IEEE international conference on Human-Robot Interaction*. ACM, 343–350.

[5] Colleen M Carpinella, Alisa B Wyman, Michael A Perez, and Steven J Stroessner. 2017. The robotic social attributes scale (rosas): Development and validation. In *Proceedings of the 2017 ACM/IEEE International Conference on human-robot interaction*. ACM, 254–262.

[6] Jeffrey T Child and Sandra Petronio. 2011. Unpacking the paradoxes of privacy in CMC relationships: The challenges of blogging and relational communication on the internet. *Computer-mediated communication in personal relationships* (2011), 21–40.

[7] Kate Darling. 2015. 'Who's Johnny?'Anthropomorphic Framing in Human-Robot Interaction, Integration, and Policy. *Anthropomorphic Framing in Human-Robot Interaction, Integration, and Policy (March 23, 2015). ROBOT ETHICS* 2 (2015).

[8] Ewart J de Visser, Samuel S Monfort, Ryan McKendrick, Melissa AB Smith, Patrick E McKnight, Frank Krueger, and Raja Parasuraman. 2016. Almost human: Anthropomorphism increases trust resilience in cognitive agents. *Journal of Experimental Psychology: Applied* 22, 3 (2016), 331.

[9] Eun Go. 2015. Does message interactivity help or hinder the effects of anthropomorphic online chat agents? Compensation vs. expectation effects in organizational websites. (2015).

[10] D. Gowda, K. Cook-Shultz, and E. Mierzwinski. 2018. Trouble In Toyland. https://uspirgedfund.org/sites/pirg/files/cpn/USN-112117-A1-REPORT/trouble-in-toyland-32.html

[11] Annabell Ho, Jeff Hancock, and Adam S Miner. 2018. Psychological, relational, and emotional effects of self-disclosure after conversations with a chatbot. *Journal of Communication* 68, 4 (2018), 712–733.

[12] Yifeng Hu and S Shyam Sundar. 2010. Effects of online health sources on credibility and behavioral intentions. *Communication research* 37, 1 (2010), 105–132.

[13] Ming-Hui Huang. 2003. Designing website attributes to induce experiential encounters. *computers in Human Behavior* 19, 4 (2003), 425–442.

[14] Jiun-Yin Jian, Ann M Bisantz, and Colin G Drury. 2000. Foundations for an empirically determined scale of trust in automated systems. *International Journal of Cognitive Ergonomics* 4, 1 (2000), 53–71.

[15] L Crystal Jiang, Natalya N Bazarova, and Jeffrey T Hancock. 2013. From perception to behavior: Disclosure reciprocity and the intensification of intimacy in computer-mediated communication. *Communication Research* 40, 1 (2013), 125–143.

[16] Malte F Jung, Jin Joo Lee, Nick DePalma, Sigurdur O Adalgeirsson, Pamela J Hinds, and Cynthia Breazeal. 2013. Engaging robots: easing complex human-robot teamwork using backchanneling. In *Proceedings of the 2013 conference on Computer supported cooperative work*. ACM, 1555–1566.

[17] Peter H Kahn Jr, Takayuki Kanda, Hiroshi Ishiguro, Brian T Gill, Jolina H Ruckert, Solace Shen, Heather E Gary, Aimee L Reichert, Nathan G Freier, and Rachel L Severson. 2012. Do people hold a humanoid robot morally accountable for the harm it causes?. In *Proceedings of the seventh annual ACM/IEEE international conference on Human-Robot Interaction*. ACM, 33–40.

[18] Ruogu Kang, Stephanie Brown, Laura Dabbish, and Sara Kiesler. 2014. Privacy attitudes of mechanical turk workers and the us public. In *10th Symposium On Usable Privacy and Security ({SOUPS} 2014)*. 37–49.

[19] Matthias U Keysermann, Henriette SM Cramer, Ruth Aylett, Carsten Zoll, Sibylle Enz, and Patrícia Amâncio Vargas. 2012. Can I trust you?: sharing information with artificial companions.. In *AAMAS*. 1197–1198.

[20] Jong-Eun Roselyn Lee and Clifford I Nass. 2010. Trust in computers: The computers-are-social-actors (CASA) paradigm and trustworthiness perception in human-computer communication. In *Trust and technology in a ubiquitous modern environment: Theoretical and methodological perspectives*. IGI Global, 1–15.

[21] Jin Joo Lee, Fei Sha, and Cynthia Breazeal. 2019. A Bayesian Theory of Mind Approach to Nonverbal Communication. In *2019 14th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*. IEEE, 487–496.

[22] SeoYoung Lee and Junho Choi. 2017. Enhancing user experience with conversational agent for movie recommendation: Effects of self-disclosure and reciprocity. *International Journal of Human-Computer Studies* 103 (2017), 95–105.

[23] Daniel T Levin, Caroline Harriott, Natalie A Paul, Tao Zhang, and Julie A Adams. 2013. Cognitive dissonance as a measure of reactions to human-robot interaction. *Journal of Human-Robot Interaction* 2, 3 (2013), 3–17.

[24] Daniel T Levin, Stephen S Killingsworth, and Megan M Saylor. 2008. Concepts about the capabilities of computers and robots: A test of the scope of adults' theory of mind. In *2008 3rd ACM/IEEE International Conference on Human-Robot Interaction (HRI)*. IEEE, 57–63.

[25] Laura H Lind, Michael F Schober, Frederick G Conrad, and Heidi Reichert. 2013. Why do survey respondents disclose more when computers ask the questions? *Public opinion quarterly* 77, 4 (2013), 888–935.

[26] Bingjie Liu and S Shyam Sundar. 2018. Should Machines Express Sympathy and Empathy? Experiments with a Health Advice Chatbot. *Cyberpsychology, Behavior, and Social Networking* 21, 10 (2018), 625–636.

[27] Max M Louwerse, Arthur C Graesser, Shulan Lu, and Heather H Mitchell. 2005. Social cues in animated conversational agents. *Applied Cognitive Psychology: The Official Journal of the Society for Applied Research in Memory and Cognition* 19, 6 (2005), 693–704.

[28] Gale M Lucas, Jonathan Gratch, Aisha King, and Louis-Philippe Morency. 2014. ItâĂŹs only a computer: Virtual humans increase willingness to disclose. *Computers in Human Behavior* 37 (2014), 94–100.

[29] Christoph Lutz and Aurelia Tamò. 2015. RoboCode-Ethicists: Privacy-friendly robots, an ethical responsibility of engineers?. In *Proceedings of the ACM Web Science Conference*. ACM, 21.

[30] Aleecia M McDonald and Lorrie Faith Cranor. 2008. The cost of reading privacy policies. *ISJLP* 4 (2008), 543.

[31] Yasunori Morishima, Hiroshi Nakajima, Scott Brave, Ryota Yamada, Heidy Maldonado, Clifford Nass, and Shigeyasu Kawaji. 2004. The role of affect and sociality in the agent-based collaborative learning system. In *Tutorial and Research Workshop on Affective Dialogue Systems*. Springer, 265–275.

[32] Clifford Nass, Jonathan Steuer, Ellen Tauber, and Heidi Reeder. 1993. Anthropomorphism, agency, and ethopoeia: computers as social actors. In *INTERACT'93 and CHI'93 conference companion on Human factors in computing systems*. ACM, 111–112.

[33] Clifford Nass, Jonathan Steuer, and Ellen R Tauber. 1994. Computers are social actors. In *Proceedings of the SIGCHI conference on Human factors in computing systems*. ACM, 72–78.

[34] Clifford Nass, Leila Takayama, and Scott Brave. 2006. Advances in Management Information Systems, chapter Socializing Consistency: From Technical Homogeneity to Human Epitome. (2006).

[35] Amy Ogan, Vincent Aleven, Julia Kim, and Christopher Jones. 2010. Developing interpersonal relationships with virtual agents through social instructional dialog. In *International Conference on Intelligent Virtual Agents*. Springer, 236–249.

[36] S Petronio. 2002. Boundaries of Privacy: Dialectics of Disclosure (State University of New York Press, Albany, NY). (2002).

[37] Sheizaf Rafaeli. 1988. Interactivity: From new media to communication. *Advancing communication science: Sage annual review of communication research* 16 (1988), 110–134.

[38] L Raine. 2016. The state of privacy in post- Snowden America. http://www.pewresearch.org/

[39] Paul Robinette, Wenchen Li, Robert Allen, Ayanna M Howard, and Alan R Wagner. 2016. Overtrust of robots in emergency evacuation scenarios. In *The Eleventh ACM/IEEE International Conference on Human Robot Interaction*. IEEE Press, 101–108.

[40] Oscar J Romero, Ran Zhao, and Justine Cassell. 2017. Cognitive-Inspired Conversational-Strategy Reasoner for Socially-Aware Agents.. In *IJCAI*. 3807–3813.

[41] Shruti Sannon and Dan Cosley. 2019. Privacy, Power, and Invisible Labor on Amazon Mechanical Turk. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. ACM, 282.

[42] S Shyam Sundar. 2008. The MAIN model: A heuristic approach to understanding technology effects on credibility. *Digital media, youth, and credibility* 73100 (2008).

[43] Dag Sverre Syrdal, Michael L Walters, Nuno Otero, Kheng Lee Koay, and Kerstin Dautenhahn. 2007. He knows when you are sleeping-privacy and the personal robot companion. In *Proc. Workshop Human Implications of Human-Robot Interaction, Association for the Advancement of Artificial Intelligence (AAAIâĂŹ07)*. 28–33.

[44] Roger Tourangeau, Mick P Couper, and Darby M Steiger. 2003. Humanizing self-administered surveys: experiments on social presence in web and IVR surveys. *Computers in Human Behavior* 19, 1 (2003), 1–24.

[45] Vivian Tsai, Timo Baumann, Florian Pecune, and Justine Cassell. 2019. Faster responses are better responses: Introducing incrementality into sociable virtual personal assistants. In *9th International Workshop on Spoken Dialogue System Technology*. Springer, 111–118.

[46] Astrid M Von der Puetten, Nicole C Krämer, Jonathan Gratch, and Sin-Hwa Kang. 2010. It doesn't matter what you are! Explaining social effects of agents and avatars. *Computers in Human Behavior* 26, 6 (2010), 1641–1650.

[47] Liz C Wang, Julie Baker, Judy A Wagner, and Kirk Wakefield. 2007. Can a retail web site be social? *Journal of marketing* 71, 3 (2007), 143–157.

[48] Yang Wang, Pedro Giovanni Leon, Alessandro Acquisti, Lorrie Faith Cranor, Alain Forget, and Norman Sadeh. 2014. A field trial of privacy nudges for facebook. In *Proceedings of the 32nd annual ACM conference on Human factors in computing systems*. ACM, 2367–2376.

[49] Susan Waters and James Ackerman. 2011. Exploring privacy management on Facebook: Motivations and perceived consequences of voluntary disclosure. *Journal of Computer-Mediated Communication* 17, 1 (2011), 101–115.

[50] Suzanne Weisband and Sara Kiesler. 1996. Self disclosure on computer forms: Meta-analysis and implications. In *Proceedings of the SIGCHI conference on human factors in computing systems*. ACM, 3–10.

[51] Heng Xu, Hock-Hai Teo, and Bernard Tan. 2005. Predicting the adoption of location-based services: the role of trust and perceived privacy risk. *ICIS 2005 proceedings* (2005), 71.

[52] Ran Zhao, Alexandros Papangelis, and Justine Cassell. 2014. Towards a dyadic computational model of rapport management for human-virtual agent interaction. In *International Conference on Intelligent Virtual Agents*. Springer, 514–527.

[53] Youyou Zhou. 2018. An Oregon family's encounter with Amazon Alexa exposes the privacy problem of smart home devices. https://qz.com/1288743/amazon-alexa-echo-spying-on-users-raises-a-data-privacy-problem/

Received June 2019; revised October 2019; accepted November 2019