Decision Support Systems 114 (2018) 94–102

Contents lists available at ScienceDirect

Decision Support Systems

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

# The influence of conversational agent embodiment and conversational relevance on socially desirable responding

Ryan M. Schuetzler[a,*], Justin Scott Giboney[b], G. Mark Grimes[c], Jay F. Nunamaker Jr.[d]

[a] University of Nebraska at Omaha, United States
[b] Brigham Young University, United States
[c] University of Houston, United States
[d] University of Arizona, United States

ARTICLE INFO

*Keywords:*
Social presence
Conversational agents
Disclosure
Social desirability

ABSTRACT

Conversational agents (CAs) are becoming an increasingly common component in a wide range of information systems. A great deal of research to date has focused on enhancing traits that make CAs more humanlike. However, few studies have examined the influence such traits have on information disclosure. This research builds on self-disclosure, social desirability, and social presence theories to explain how CA anthropomorphism affects disclosure of personally sensitive information. Taken together, these theories suggest that as CAs become more humanlike, the social desirability of user responses will increase. In this study, we use a laboratory experiment to examine the influence of two elements of CA design—conversational relevance and embodiment—on the answers people give in response to sensitive and non-sensitive questions. We compare the responses given to various CAs to those given in a face-to-face interview and an online survey. The results show that for sensitive questions, CAs with better conversational abilities elicit more socially desirable responses from participants, with a less significant effect found for embodiment. These results suggest that for applications where eliciting honest answers to sensitive questions is important, CAs that are "better" in terms of humanlike realism may not be better for eliciting truthful responses to sensitive questions.

## 1. Introduction

Advances in technology since the mid-1990s have ushered in a new age of communication where many face-to-face (FtF) interactions have been replaced by interactions between humans and computers. These interactions may be in the form of computer mediated communication between two or more humans, or in the form of human-computer interactions, in which the computer is the ultimate communication partner. While many human-computer interactions remain clearly in the domain of a human interacting with a computer using conventional methods and norms (i.e., using the keyboard or mouse to perform specific tasks), an emerging area of interest is the replacement of human agents with *conversational agents* (CAs)—systems that mimic human-to-human communication using natural language processing, machine learning, and/or artificial intelligence [1].

The idea of interacting with a computer as if it were another human has fascinated users and developers of information systems for many years. Early implementations of CAs were novelties designed to play specific roles such as the Rogerian psychotherapist ELIZA [2], and PARRY—a paranoid patient [3]. As technological capabilities have advanced, these "toy" CAs have given way to the emergence of sophisticated and generalizable frameworks that parse user responses and mimic understanding by responding to pre-defined phrases or keywords (e.g., A.L.I.C.E. [4] and ChatScript [5]) [6]. These and other similar platforms have recently ignited a substantial increase in the popularity of CAs and many popular instant messaging and social media platforms, such as Facebook Messenger, WhatsApp, and Kik, have integrated tools to develop and deploy CAs. These efforts have been met with enthusiastic response from users. For example, in the year following the introduction of its bot integration platform in 2016, Facebook Messenger saw the introduction of over 34,000 conversation agents, or "bots" [7].

This increase in pervasiveness and utility has resulted in CAs taking on more serious roles such as serving as virtual personal assistants [8], conducting medical interviews [9,10], providing therapy for depression and anxiety [11], disseminating emergency response information [12],

* Corresponding author.
*E-mail addresses:* rschuetzler@unomaha.edu (R.M. Schuetzler), justin_giboney@byu.edu (J.S. Giboney), gmgrimes@bauer.uh.edu (G.M. Grimes), jnunamaker@cmi.arizona.edu (J.F. Nunamaker).

https://doi.org/10.1016/j.dss.2018.08.011
Received 22 May 2018; Received in revised form 20 July 2018; Accepted 28 August 2018
Available online 29 August 2018

and conducting interviews to detect fraud and deception [13,14]. In many of these scenarios, the information being solicited may be considered sensitive and individuals may be unwilling or hesitant to disclose the information—not necessarily for nefarious reasons, but rather to avoid providing answers society would deem unacceptable or confessing undesirable behavior [15]. Because of the wide variety of contexts in which CAs operate, understanding how specific design choices influence user perceptions and behaviors is an important topic of study.

While prior research has thoroughly explored the mechanics of using CAs to conduct interviews and how to make CAs more humanlike, only recently has attention been paid to how design decisions may impact how comfortable users are disclosing potentially sensitive information to a CA [16]. It has been suggested that CAs that are perceived as more humanlike may have the unintended consequence of increasing discomfort in users [17,18]. As emerging applications are using CAs to elicit sensitive information from users—for example, in a medical office performing the interviewing duties of an intake nurse [9]—it is important to understand the effect more humanlike CAs have on information disclosure. The way a question is asked, and who is doing the asking, can have strong effects on the truthfulness of answers given [19,20]. Thus, such design decisions are critical when sensitive personal information must be elicited.

In pursuit of empirically studying the effect of making a CA more humanlike on disclosure of sensitive information, this paper builds on self-disclosure, social desirability, and social presence research. We examine how people adapt the social desirability of their answers in response to the social presence of a CA interviewer, compared to an online survey and a face-to-face interview. The following research question guides this work:

*How do the conversational capabilities and embodiment of a CA influence disclosure of sensitive information?*

## 2. Theory and hypotheses

### 2.1. Social presence

The influence of humanlike characteristics, such as the capability to hold a conversation and representative embodiment, is explained by *social presence*—the sense of connection that a user feels with their communication partner [21]. Social presence is frequently manipulated via attributes of the communication medium, such as its richness [22,23]. As a communication medium allows for richer content, the media evokes a greater sense of social presence compared to less rich media [24] and can give additional context to communication [25], thus increasing social presence. In addition to the richness of the medium, the way in which the medium is used and the information conveyed—i.e., the conversational capability of one's partner—also influence perceptions of social presence [26,27].

Given our understanding of how users perceive computers as social actors [28], the influence of social presence on disclosure should apply whether the conversation partner is a human or a computer. Prior research has found that people often treat computer systems as if they were human [1,29]—for example, by applying politeness norms [30], reciprocating self-disclosure [31], and expressing a feeling of connection [32]. In the case of information disclosure, social presence could have either positive or negative effects. On the positive side, social presence can increase trust [33], potentially making people feel more comfortable disclosing. Conversely, greater social presence can also result in negative outcomes as people consider the social desirability of their responses and how their responses might influence their communication partner's opinion of them [34]. We suggest that in an interview situation, particularly one in which sensitive information is being elicited, a greater sense of social presence will evoke more socially desirable responses, in which people are more likely to adjust their responses to match what they think the socially desirable response is.

### 2.2. Self-disclosure and social desirability

For many emerging applications, a core component of enhancing the usefulness of the system is encouraging users to provide information about themselves to the system. When soliciting sensitive information, the effects of attributes of the interviewer on self-disclosure and social desirability must be considered. *Self-disclosure* is the extent to which individuals share information about themselves purposely and voluntarily [35,36]. Information being disclosed about oneself may present the discloser in a positive, negative, or neutral way, and questions may or may not be viewed by the discloser as being sensitive [37]. With this in mind, a respondent may choose to disclose more or less information—or not disclose any information at all—based on the nature of the interaction.

In addition to deciding how much information to disclose, people may also modify their response to questions to increase the *social desirability* of their response. Social desirability describes the way in which people would like to be seen by others [38]. Modifying responses to be more socially desirable may stem from a desire to improve social status, or to avoid negative consequences. When people are asked to disclose socially undesirable information about themselves social desirability bias can have a strong effect on reporting [39].

Prior research has found the level of social presence in the way questions are administered can result in important differences in responses. Interactions with lower social presence, such as computer-administered surveys, have been found to result in responses that are less biased by social desirability than those in face-to-face interviews [20]. The effect of social desirability in survey responses has been studied extensively, as it presents a serious threat to the validity of survey measures [19,40,41]. Techniques such as indirect questioning [42] and self- and computer-administration of surveys [43], as opposed to human interviewing, are often used to mitigate the effects of social desirability. In line with these findings, we expect that respondents will vary the social desirability of their response in accord with the social presence of the interview format. We hypothesize that in the format with the highest social presence (face-to-face) the amount of information disclosed will be the least, and as social presence is reduced—from face-to-face to interaction with a CA, and finally to a non-interactive survey—the level of disclosure will increase. Thus we propose H1:

**H1.** Interview modalities with higher social presence lead to more socially desirable responding.

When asking interview questions, one important consideration is the sensitivity of the questions being asked, as sensitive questions are more likely to be influenced by social desirability than non-sensitive questions [44]. Among the general population, questions about topics such as medical history, sexual history, and drug/alcohol use are typically considered sensitive [37]. Sensitive questions may result in either nonresponse or high measurement error compared to non-sensitive questions [45], and may elicit less truthful responses as answering them truthfully may cause negative consequences such as shame or punishment [46]. While the aforementioned topics are generally considered to be sensitive, the sensitivity of specific questions is dependent on the individual being asked the question, the asker of the question, and the social acceptability of the topic [45]. Sensitivity can be measured through nonresponse on survey items, or through separate ratings from people indicating their willingness to answer truthfully [47,48].

Since sensitivity and social desirability depend on both the individual and the context, the same question may be of different levels of sensitivity and social desirability for different people, or even for the same person in different circumstances, thus leading to different levels of disclosure [45]. For example, individuals who are under the legal age to consume alcohol tend to overestimate drinking behaviors of their peers, potentially increasing the perceived desirability of this behavior within that group [49]. Therefore, if a person that is under the legal age to drink alcohol is asked about drinking behavior by a peer, the