The Influence of Conversational Agents on Socially Desirable Responding

Ryan M. Schuetzler
University of Nebraska at Omaha, ryan.schuetzler@byu.edu

G. Mark Grimes
University of Houston, gmgrimes@bauer.uh.edu

Justin Scott Giboney
Brigham Young University, justin_giboney@byu.edu

Jay F. Nunamaker Jr.
University of Arizona, jnunamaker@cmi.arizona.edu

The Influence of Conversational Agents on Socially Desirable Responding

Ryan M. Schuetzler
University of Nebraska at Omaha
rschuetzler@unomaha.edu

Justin Scott Giboney
Brigham Young University
justin_giboney@byu.edu

G. Mark Grimes
University of Houston
gmgrimes@bauer.uh.edu

Jay F. Nunamaker, Jr.
University of Arizona
jnunamaker@cmi.arizona.edu

Abstract

Conversational agents (CAs) are becoming an increasingly common component in many information systems. The ubiquity of CAs in cell phones, entertainment systems, and messaging applications has led to a growing need to understand how design choices made when developing CAs influence user interactions. In this study, we explore the use case of CAs that gather potentially sensitive information from people—for example, in a medical interview. Using a laboratory experiment, we examine the influence of CA responsiveness and embodiment on the answers people give in response to sensitive and non-sensitive questions. The results show that for sensitive questions, the responsiveness of the CA increased the social desirability of the responses given by participants.

1. Introduction

Advances in technology since the mid-1990s have ushered in a new age of communication where many face-to-face interactions have been replaced by interactions between humans and computers.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 the computer may act as one of the participants in the communication [21].

In many of these emerging human-computer interactions, *conversational agents* (CAs)—user interfaces that emulate human-to-human communication using natural language processing, machine learning, and artificial intelligence—are a core component of the interaction. CAs are becoming increasingly common in our everyday lives in a wide variety of contexts including virtual assistants like Siri, medical interviews [3], therapy for depression and anxiety [7], and assistance for the cognitively impaired [45]. 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. A great deal of research and development has focused on improving the social presence of CAs by enhancing the responsiveness—the ability of the agent to provide responses contingent on user messages—and embodiment—the visual representation of the agent. However, overly realistic CAs may have the unintended consequence of increasing discomfort in users [18, 36].

Of particular interest in this regard is the emerging use case of CAs soliciting sensitive personal information, for example, in a medical office performing the interviewing duties of an intake nurse. In such a case, the CA is used to gather information from patients about their medical history and reason for their visit. In order for these CAs to be effective, they must be designed in such a way that patients disclose sensitive information to them.

Though researchers have given great thought to the mechanics of using CAs to gather data, little attention has been paid to how design decisions may impact disclosure behavior. These design decisions are most critical when personal information must be elicited because people guard that information more carefully. As we move toward a world with more communication with CAs it is important that we understand how users interact and perceive the CA experience, and how CA characteristics can affect that experience. To this end, the focus of this research is to understand how responsiveness and embodiment of a CA lead human interactants to disclose or not disclose sensitive information. Thus, the following research question guides the testing of these effects:

> *How do CA characteristics influence user responses when discussing sensitive information?*

To study disclosure behavior, this paper builds on social desirability research, examining how people adapt the social desirability of their answers in response to system attributes that give a CA interaction more or less social presence. In an experiment, where participants may disclose varyingly sensitive

HICSS

information, we show that while the responsiveness of a CA increases socially desirable responding for sensitive questions, the same effect is not present for non-sensitive questions.

2. Theory and hypotheses

Conversational agents have a long and rich history in the world of information systems, with CAs taking a wide variety of forms and personalities. Early CAs were designed to play specific roles such as ELIZA, a Rogerian psychotherapist [43], or PARRY, a paranoid patient [4]. In the following decades, more advanced and generalizable frameworks have emerged such as A.L.I.C.E. [41] that parse user responses and mimic understanding by responding to certain phrases or key words. Recent advances in CAs and the explosion of popular instant messaging applications like Facebook Messenger, WhatsApp, and Kik have led to increases in the use of CAs. For example, just a year after announcing its bot integration platform, Facebook Messenger has seen the introduction of over 34,000 conversation agents, or "bots" [24].

In addition to the applications of conversational technology listed above, recent research has also endorsed the use of CAs in conducting interviews for fraud [28] and deception detection [23]—both scenarios in which individuals may be unwilling or hesitant to disclose information. Similarly, patients in healthcare settings are also often unwilling to disclose information—not for nefarious reasons, but rather to maintain face. They know the answer society would deem acceptable and do not want to confess to undesirable behavior [14]. In the healthcare domain, CAs have been investigated for their usefulness in gathering patient information and diagnosing illness [3].

2.1. Disclosure and social desirability

Self-disclosure is the extent to which individuals share information about themselves purposely and voluntarily [25, 30]. Information being disclosed may be positive, negative, or neutral, and some questions may be viewed by the respondent as sensitive [38]. To this end, the respondent may manage their disclosure depending on the nature and sensitivity of the question. Question sensitivity is dependent on the individual being asked the question, the asker of the question, and the social acceptability of the subject [39]. The same question may be of different levels of sensitivity for different people, or even for the same person in different circumstances.

People may modify how they respond to such questions in order to increase the *social desirability* of their response. Social desirability describes the tendency of people to answer questions in such a way as to present themselves in the best light [5]. Social desirability can be particularly influential when asking sensitive questions, such as those involving health behaviors, sexual history, drug use, or alcohol consumption [38]. These questions are considered sensitive since answering them truthfully may cause negative consequences such as shame, and if the answers were disclosed or passed to the wrong entity they could have a negative impact on the discloser's relationships or even career. As such, these questions are most likely to be influenced by social desirability.

For example, underage individuals tend to overestimate drinking behaviors of their peers, potentially increasing the perceived desirability of this behavior within that group [26]. Therefore, if a person that is under the legal age to drink alcohol is asked about drinking behavior by a peer, the question might be considered positive and of low sensitivity. Thus, the respondent would be willing to disclose, and perhaps even inflate, their drinking behavior to improve the social desirability of their response. However, if an authority figure asks the same individual about drinking, the question may be deemed sensitive and of negative valence, thus leading the respondent to hide or underreport drinking to avoid punishment [6]. For someone who is of drinking age or someone that does not drink, this question may be of low sensitivity and have a neutral valence. However, even for someone of drinking age, if they feel their drinking is outside of what is deemed socially acceptable, the question may have heightened sensitivity and negative valence. Similar behaviors have been found with exercise and consumption of healthy foods [1], albeit to a lesser extent [11].

Prior research has found the method of question administration can lead to important differences in responses. For example, computer-administered surveys generally result in responses that are less biased by social desirability than those in face-to-face interviews [32]. The effect of social desirability has been studied extensively in surveys, as it presents a serious threat to the validity of survey measures [12, 13, 16]. To gather accurate survey data when social desirability is a concern, steps must be taken to measure and/or minimize its effect [22]. Indirect questioning [22] and self- and computer-administration of surveys (as compared to human interviewing) [37] are methods used to mitigate the effects of social desirability. Because of the effects of social desirability, it is important to explore how more humanlike CAs affect disclosure. The influence of these humanlike characteristics can be explained by social presence.

2.2. Social presence

Social presence is a sense of connection that a user feels with their communication partner [34]. Whether that communication partner is a computer system or another human, varying aspects of the communication medium can influence the degree to which people are thinking about what is on the other side of the communication [8]. In the case of information disclosure, social presence could have either positive or negative effects. On the positive side, social presence can increase trust [8], 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 [40].

The Computers Are Social Actors (CASA) paradigm suggests that in many ways, people will treat a computer system as if it were a person [20, 21]. Through dozens of studies, researchers have shown that people respond to computer systems in similar ways to how they would respond to a human—for example, by applying politeness norms [31], reciprocating self-disclosure [17], and attributing personality to computer partners [20]. These findings, to a degree, contradict what would be suggested by the social desirability literature supporting computer-administration of surveys—if computers are attributed personality, why are they better at extracting undesirable opinions or responses? We suggest this may be a function of the level of anthropomorphism projected upon the computer by the user.

To investigate this, previous studies have used CAs to accompany survey administration [13, 15], but investigations to this point have not examined the social desirability effects of giving the CAs conversational responsiveness, and embodiment. Conversational *responsiveness* of a CA refers to the ability of the agent to provide the appearance of understanding the user's input by responding in a contingent manner. To illustrate, consider a CA that asked a user the question, "What is your favorite movie?" There are a multitude of responses the user could provide. A non-responsive CA will provide a generic response regardless of the answer provided by the user, while a responsive CA will parse the user's message and give a response that is related to the content. For example, if the user responds with "Saving Private Ryan," the CA might respond with "I don't watch many war movies." Likewise, if the user responds with "The Notebook," the CA might reply "Can't go wrong with Nicholas Sparks." This type of contingent reply can give the impression that the CA understood the input, thus mimicking human-to-human conversation and creating a more natural conversational flow.

CAs that communicate well are perceived to understand the user and can therefore make judgments about their responses [29]. This capability has been shown to increase anxiety in social-phobic patients—those who fear interacting with and being evaluated by other people—when engaging with highly interactive CAs [9]. The level of (dis)comfort that is present during the solicitation of information can have a significant effect on the accuracy of the information. Computer-assisted self-administration of surveys can lead to more accurate responses to potentially embarrassing questions [38]. However, humanizing the computer system with responsiveness and embodiment might negate the benefits of the computer-based system [37]. Therefore, we hypothesize:

H1: Increasing agent conversational responsiveness will increase participants' socially desirable responding.

Another form of social presence manipulation is embodiment. *Embodiment* refers to an agent having a visual representation. Research on embodiment effects on social presence have often used avatars—digital representations of real people—rather than automated agents [2]. When a CA is given an animated facial representation, it makes the CA appear more human-like [10, 35] and increases the naturalness of the communication [35]. Therefore, a CA with an animated facial representation will have higher social presence.

Prior research suggests that the mere presence of a face in human and computer-administered surveys creates pressure to respond in socially desirable ways [13]. Examples in human interviews include the confessional booth or a psychoanalyst's couch, both situations where the interviewer's face is hidden from the discloser with the intent of encouraging more candid responses. Lind et al. [13] showed a strong effect of facial representation on socially desirable responding to surveys, with people showing more socially desirable responding when interacting with a face than with text alone. Therefore, we hypothesize:

H2: Adding agent visual embodiment will increase socially desirable responding.

When asking interview questions, one important consideration is the sensitivity of the questions being asked. Sensitive information, for example that relating to financial or medical conditions, is more likely to influence the way people respond [11]. When people are asked to disclose socially undesirable information about themselves, such as poor academic performance or embarrassing medical conditions, social desirability bias can have a strong effect on reporting [33]. Such

effects may be seen less frequently, or not at all, when reporting socially desirable information such as excellent academic performance or positive exercise behaviors [11]. The sensitivity of a question will vary between individuals depending on the person's behavior [39]. In the case of asking about smoking, someone who never smoked should have no problem, an ex-smoker might be very sensitive, a closet smoked might be very sensitive, a heavy unapologetic smoked might not care, and a heavy smoker who knows it's bad might be sensitive. While the sensitivity of a question is not universal, there are still patterns in the types of questions judged to be sensitive. Among the general population, survey questions about topics such as substance abuse, political behavior, and income are frequently considered sensitive, and result in either nonresponse or high measurement error compared to non-sensitive topics [39]. This research on the effects of question sensitivity leads us to the following hypothesis regarding the moderating effect of question sensitivity:

H3: The influence of conversational agents on socially desirable responding will exist only for sensitive questions.

3. Method

To test these hypotheses, we employed a 2x2 laboratory experiment. The sample and method of the experiment are described here.

3.1. Sample

One hundred and twenty-nine native English speaking participants were recruited from an introductory management information systems course at a large U.S. university. Participants received course credit and a small monetary compensation for their time. Of the 129 participants, 15 failed the attention checks built into the study, leaving 114 participants (59 female). The average age of the participants was 20.9 years with a standard deviation of 1.9 years.

3.2. Conversational agent design

We used the ChatScript engine to create the interviewing agent for this study [44]. The conversational agent creation process involved three main steps. First, we chose conversation topics relevant to the subject pool. For example, since the participants were college students, the agent asked about their major, classes, and favorite activities in the area. Next, a corpus of patterns and anticipated answers to questions related to these topics was created. For example, on the topic of majors, if the participant reported computer science as their major, the CA would respond with a message such as, "That's cool, I love technology."

Using this initial conversation corpus, we conducted a pilot test to identify potential responses for which matching patterns did not exist. While it is infeasible to match every possible response a user might give, due to the limited scope of the conversation topics we were able to create responses for the majority of inputs given by participants. Following the pilot test, we created new patterns for any non-matched utterances.

3.3. Procedure

To identify topics of varying levels of sensitivity, we first created a list of potential interview questions identified as sensitive or nonsensitive topics by prior research [13]. As part of a separate data collection from the same population as the main study, we asked participants to rate from 1 to 6 how comfortable they would feel answering specific questions about each topic. Among the topics considered, the largest difference in sensitivity was between health and drinking behaviors. Two corresponding questions from each topic were chosen to represent these topics (Table 1). To avoid violating normality assumptions due to the data being skewed, we used a paired Wilcox signed rank test to evaluate the differences in sensitivity between topics. Health (M=5.16) and drinking (M=4.67) behavior were found to be statistically different (n=138, V=2070.5, p<.001).

Table 1. Interview questions

Drinking Behavior (high sensitivity)	How many alcoholic drinks do you have in a typical week?
	How many times in the past 30 days did you drink to the point of intoxication?
	How many total servings of fruit and/or vegetables did you eat yesterday?
	On how many of the past 7 days did you exercise for at least 20 minutes?

We used a 2 (responsive vs. nonresponsive) x 2 (embodied vs. unembodied) between-subjects experimental design to test the hypotheses. In the nonresponsive condition, the CA provided the same response regardless of user input. In the responsive condition the CA conversed more naturally by engaging in follow-up conversation based on matching user input in the conversation corpus. The different agents used the same number of utterances in both interviews to ensure that users saw and answered the same number of

questions. In the unembodied condition, the chat took place without a visual avatar. In the embodied avatar condition, participants interacted with a CA that had an animated face.
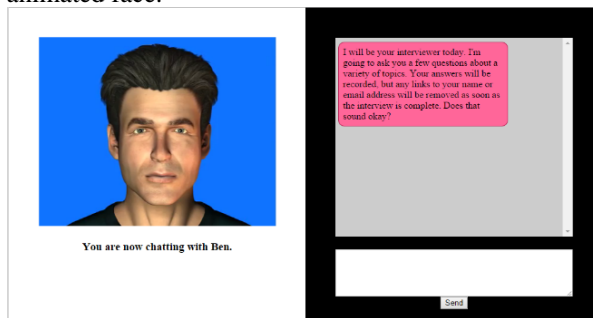

Figure 1. Chat interface with visually embodied agent

After completing a pre-experiment survey, participants reported to a computer lab containing 25 computers. To preclude the possibility of contamination due to a participant in the unembodied condition seeing the embodied agent on a nearby computer screen, each session was randomly assigned to an embodied or unembodied agent condition. Participants were randomly assigned in real-time to either the responsive or nonresponsive condition, as these two conditions appear identical at a glance.

Each chat interview began with basic questions identified as rapport-building by Lucas et al. [15]. These questions include general introductory questions such as "What class are you here for?" and "What is your favorite outdoor activity?" It is during these introductory questions that the majority of the differences between the responsive and nonresponsive CAs were introduced. The nonresponsive CA gave generic follow-up questions to each response. For example, the question about outdoor activities was followed with "What else do you enjoy doing?" The responsive bot, on the other hand, gave different responses based on the user's response. If the user responded with "swimming," the bot would follow up with "Water sports are fun. How often do you go?" Similarly, if the user instead said "hiking," the bot responded with "I've wanted to try hiking for a while now. When did you start?" A wide variety of responses were matched in this way to create a conversational tone for the interview. For both the responsive and non-responsive bots, the number of questions asked and the approximate number of words shown was kept consistent across conditions. After the rapport-building questions, the CA asked the previously described interview questions (see Table 1).

After the interview, participants were directed to a post survey where demographic information was collected and follow-up questions about the interview

were posed to the participants. Specifically, participants were asked to rate how truthful they had been with the answers they gave in the interview on a 5-point, Likert-like scale. After completing the survey participants received them remuneration and were free to leave.

4. Analysis

We expected that participants would give more socially desirable responses when interacting with more human-like CAs. To group each topic, we to first standardized the responses for each question and then averaged the topic responses for each participant. We tested the hypotheses using two separate generalized linear models; one for each topic (see Table 2). We controlled for age and sex.

H1 predicted that participants interacting with a responsive CA will give more socially desirable answers. For drinking behavior, the GLM reports a significant direct effect for responsiveness, controlling for gender. Participants in the responsive condition reported, on average, less excessive drinking behavior than participants in the non-responsive condition. Thus, H1 was supported.

H2 predicted that participants interacting with an embodied CA will give more socially desirable answers. For drinking behavior, the GLM does not report a direct effect for visual embodiment. Thus, H2 was not supported.

We also examined the interaction between responsiveness and visual embodiment for drinking behavior. Figure 2 shows the interaction between the two conditions. While the GLM does not report a significant result, there does seem to be a possible interaction. We will elaborate on this in the discussion.

H3 predicted that the effects of responsiveness and embodiment will only be significant in more sensitive questions. We compare the GLM between drinking behavior and health behavior and we note that none of the conditions are significant for health behavior. Thus, H3 is supported.

We also examined the interaction between responsiveness and visual embodiment for drinking behavior. Figure 3 shows the interaction between the two conditions. While the GLM does not report a significant result, there does seem to be a possible interaction. We will elaborate on this in the discussion. Table 3 presents a concise summary of the results.
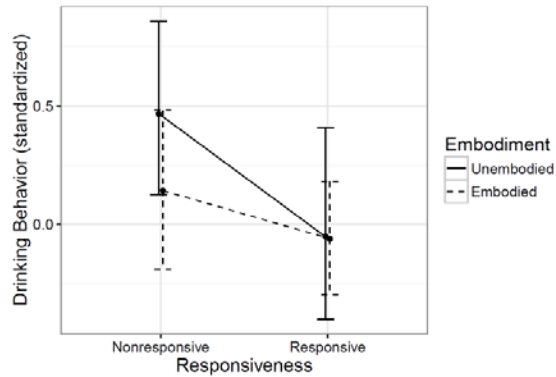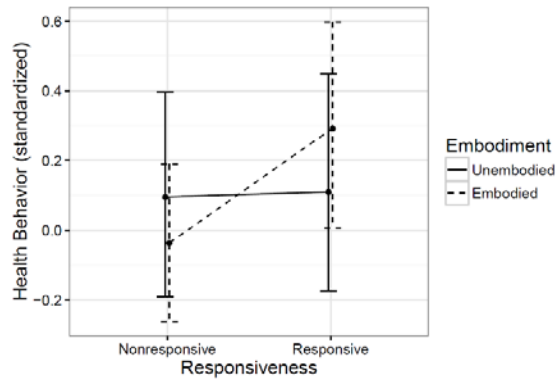
Figure 2. Drinking behavior (standardized)

Figure 3. Health behavior (standardized)

5. Discussion

For this study, we developed a web-based chat interface and a CA to interact with users. In our 2x2 experiment design, participants interacted with either a responsive or a nonresponsive CA that either had or did not have a visual embodiment. During the interaction, participants were asked about issues that are generally considered to be either sensitive (alcohol consumption) or not sensitive (general health behavior). Both the responsive and nonresponsive CAs asked the same initial questions. The nonresponsive Ca gave little feedback and asked generic follow-up questions while the responsive CA responded with dynamic content relevant to the answer given by the participant. We tested the relationship of this responsiveness manipulation, as well as an embodiment manipulation, on users' disclosure when responding to the system.

Table 2. Results: Means and SDs

	Drinking Behavior	Health Behavior
Intercept	0.91 (0.97)	1.01 (0.83)
Age	-0.03 (0.05)	-0.05 (0.04)
Sex	0.43* (0.18)	0.12 (0.15)
Responsiveness	-0.50* (0.25)	0.05 (0.21)
Face	-0.33 (0.25)	-0.13 (0.20)
Responsive x Face	0.35 (0.35)	0.29 (0.30)

* $p < .05$

Table 3. Results of hypothesis testing

	Hypothesis	Support
H1	Increasing responsive conversation will increase socially desirable responding.	Partial
H2	Embodiment will increase socially desirable responding.	No
H3	The influence of conversational agents on social desirability will exist only for sensitive questions.	Yes

Through our hypothesis testing we found that the responsiveness of the CA does in fact influence disclosure when interacting with a system that asks sensitive questions. Because our participants were randomly assigned to a condition, we assume that there were no systematic differences between groups in their *actual* alcohol use. However, there were statistically significant differences between the groups in their *reported* alcohol use. Our analysis used a standardized composite measure of alcohol use composed of the two alcohol-related questions from the interview. The unstandardized data show that those in the nonresponsive condition reported an average of 5.9 drinks per week ($SD = 6.4$), while those in the responsive condition reported only 4.4 drinks ($SD = 6.6$). Similarly, those in the responsive condition reported being intoxicated 2.6 days ($SD = 3.0$) in the last month, while those in the nonresponsive reported 4.2 days ($SD = 4.8$). Even though there are high individual differences in reported and actual drinking behavior, there is a clear social desirability effect, with those in the responsive condition reporting more responsible behavior. The embodiment manipulation was shown to have no statistically significant effect. Embodiment is a complex manipulation which may be influenced by many factors including the quality of the animation, the perceived social status of the avatar, gender differences [23], demeanor of the avatar, similarity to the participant

[27], and more. Future research should investigate other manipulations of embodiment.

Because of the college student population studied here, it makes sense that alcohol use is a particularly sensitive topic. Particularly among undergraduates, alcohol use and abuse are salient topics [42]. However, binge drinking is much less likely after college [19], so future research on other populations may consider different types of questions may reach the sensitivity required to induce socially desirable responding. Questions about ethnicity or income may serve as a better basis for sensitive questions [29].

We saw much smaller differences in reporting for the nonsensitive questions about healthy behavior. Individuals in the responsive condition reported eating 2.4 servings of vegetables ($SD = 1.6$) the previous day, while those with the nonresponsive reported 2.0 servings ($SD = 1.3$). There was virtually no difference at all between conditions in reported exercise in the last week (responsive: $M = 3.5$, $SD = 1.8$; nonresponsive: $M = 3.4$, $SD = 2.1$). This is consistent with previous research showing that social desirability effects of question administration mode are stronger for undesirable rather than for desirable actions [11].

We also performed a *post hoc* analysis on the participants' self-reported truthfulness scores. After the interaction with the CA, we asked participants how truthful they were on each of the questions asked during the interview. The participants' self-reported truthfulness scores could be considered another measure of self-disclosure during the interview. We ran two separate GLMs to test the effects of our conditions on sensitive and nonsensitive questions (see Table 4). This time the interaction of responsiveness and visual embodiment was significant for drinking behavior but not health behavior. Figures 4 and 5 indicate that participants reported being less truthful to a responsive CA with a visual embodiment than to all other combinations. This self-reported truthfulness score provides support for the theoretical explanation that an agent with higher social presence will produce greater socially desirable responding for sensitive questions. It also provides support that the differences in reported drinking behavior were not due to random systematic differences between groups.

When the topic of interest is one that may be sensitive to those responding, one must be careful about inducing social desirability effects with agents that might be perceived as more socially present. Even small differences like adding a minimal amount of responsiveness in communication can have significant effects on the quality of information gathered in an interview.

Table 4. Truthfulness results: Means and SDs

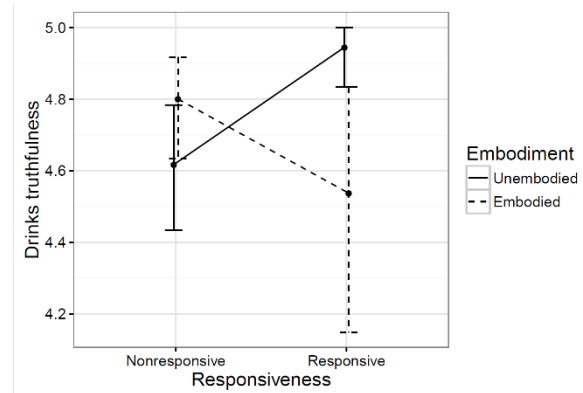|	Drinking Behavior	Health Behavior
Intercept	4.44*** (0.61)	4.39 (0.64)
Age	0.01 (0.03)	0.01 (0.03)
Sex	-0.001 (0.11)	0.09 (0.11)
Responsiveness	0.32* (0.15)	0.09 (0.16)
Face	0.18 (0.15)	0.06 (0.15)
Responsive x Face	-0.58** (0.22)	-0.36 (0.23)

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

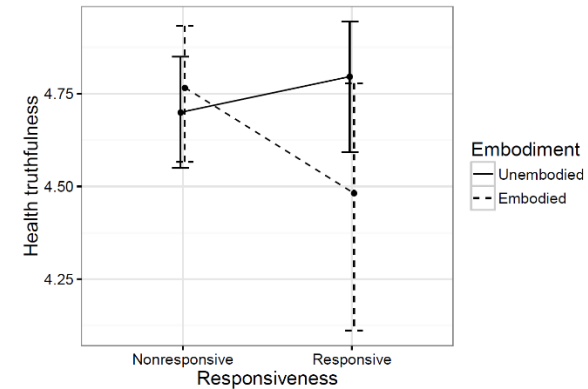**Figure 4. Drinking truthfulness**

Figure 5. Health truthfulness

5.1. Limitations

One limitation in the current study, which is common in many self-report studies, is that we cannot determine which interview scenario elicited responses that were more truthful. People who interacted with a non-responsive agent disclosed more potentially negative sensitive information, but it is impossible to say if those people were inflating the truth, accurately

reporting, or continuing to under-report. However, as there is a significant difference in the level of information disclosed, it is likely that the experimental manipulations were the cause of the difference.

Another limitation is that our study was limited to one type of questions from the nonsensitive and sensitive categories. Further studies will need to be done to ensure our results will generalize to other categories of sensitive and nonsensitive questions.

5.2. Implications for research

This work contributes to research on CAs by furthering our understanding of the benefits and potential limitations of using CAs to gather sensitive information. The current study demonstrates that the design of an interview experience influences the level of disclosure. There are other avenues to explore to understand how responsive intelligent agents can shape interactions and manipulate individual responses. Potential ideas include validating responses, empathizing, having a CA disclose embarrassing information, or manipulating the embodiment to look either less or more threatening.

5.3. Implications for practice

The results indicate that when practitioners design CAs to elicit sensitive information, they should avoid features that make the CA's interaction too human-like. As CAs are now being used in sensitive situations such as treating depression and anxiety, it is important to evaluate and consider the effects of responsiveness [7]. Each scenario or application of CAs likely has its own goals, creating different considerations for design. Creating CAs that appear more responsive leads interviewees to manage their disclosure more carefully, possibly leading them to hide embarrassing but potentially important information. For those who review the information gathered by CAs (such as doctors or prospective employers), more honest disclosure is better. When interviewees disclose more information, decision-makers can better evaluate risks. When disclosure is low, decision-makers must determine if the failure to disclose was for lack of information to disclose, or embarrassment.

The results of this study suggest that, in situations where sensitive information increases the risk of socially desirable responding, care must be taken to ensure that the information gathered is accurate. Increasing the sense of social presence in a system does not have a universally positive effect on system outcomes. While the measures in themselves may seem desirable, they can lead to negative consequences.

6. Conclusion

This study set out to investigate how CA design influences user responses and behaviors when discussing sensitive information. The study shows that CAs that are more responsive decrease the amount of sensitive information that people disclose. Therefore, this study demonstrates that being more humanlike may be a detriment to CAs intended to collect sensitive information.

7. Acknowledgements

8. References

[1] Adams, S.A., Matthews, C.E., Ebbeling, C.B., et al. The Effect of Social Desirability and Social Approval on Self-Reports of Physical Activity. *American Journal of Epidemiology 161*, 4 (2005), 389.

[2] Bente, G., Rüggenberg, S., Krämer, N.C., Eschenburg, F., Bente, G., and Ru, S. Avatar-Mediated Networking: Increasing Social Presence and Interpersonal Trust in Net-Based Collaborations. *Human Communication Research 34*, 2 (2008), 287–318.

[3] Bottles, K. Will Patients Trust Sociable Humanoid Robots. 2011. http://www.kevinmd.com/blog/2011/08/patients-trust-sociable-humanoid-robots.html.

[4] Colby, K.M., Franklin, D.H., Weber, S., and Kraemer, H. Turing-like Insdistinguishability Tests for the Validation of a Computer Simulation of Paranoid Processes. *Artificial Intelligence 3*, (1972), 199–221.

[5] Crowne, D.P. and Marlowe, D. A New Scale of Social Desirability Independent of Psychopathology. *Journal of Consulting Psychology 24*, 4 (1960), 349–54.

[6] Davis, C.G., Thake, J., and Vilhena, N. Social Desirability Biases in Self-reported Alcohol Consumption and Harms. *Addictive Behaviors 35*, 4 (2010), 302–311.

[7] Fitzpatrick, K.K., Darcy, A., and Vierhile, M. Delivering Cognitive Behavior Therapy to Young Adults With Symptoms of Depression and Anxiety Using a Fully Automated Conversational Agent (Woebot): A Randomized Controlled Trial. *JMIR Ment Health 4*, 2 (2017), e19.

[8] Gefen, D. and Straub, D.W. Consumer trust in B2C e-Commerce and the importance of social presence: experiments in e-Products and e-Services. *Omega 32*, 6 (2004), 407–424.

[9] ter Heijden, N., Qu, C., Wiggers, P., and Brinkman, W.-P. Developing a Dialogue Editor to Script Interaction Between Virtual Chracters and Social Phobic Patients.

Proceedings of the ECCE2010 Workshop: Cognitive Engineering for Technology in Mental Health Care and Rehabilitation, (2010), 111–123.

[10] Holtgraves, T.M., Ross, S.J., Weywadt, C.R., and Han, T.L. Perceiving artificial social agents. *Computers in Human Behavior 23*, 5 (2007), 2163–2174.

[11] Kreuter, F., Presser, S., and Tourangeau, R. Social Desirability Bias in CATI, IVR, and Web Surveys: The Effects of Mode and Question Sensitivity. *Public Opinion Quarterly 72*, 5 (2008), 847–865.

[12] Krysan, M. and Couper, M.P. Race in the Live and the Virtual Interview: Racial Deference, Social Desirability, and Activation Effects in Attitude Surveys. *Social Psychology Quarterly 66*, 4 (2003), 364–383.

[13] Lind, L.H., Schober, M.F., Conrad, F.G., and Reichert, H. Why Do Survey Respondents Disclose More When Computers Ask the Questions? *Public Opinion Quarterly 77*, 4 (2013), 888–935.

[14] Linehan, M.M. and Nielsen, S.L. Assessment of Suicidal Ideation and Parasuicide. *Journal of Consulting and Clinical Psychology 49*, (1981), 773–775.

[15] Lucas, G.M., Gratch, J., King, A., and Morency, L.-P. It's Only a Computer: Virtual Humans Increase Willingness to Disclose. *Computers in Human Behavior 37*, (2014), 94–100.

[16] McCrae, R.R. and Costa, P.T. Social Desirability Scales: More Substance Than Style. *Journal of Consulting and Clinical Psychology 51*, 6 (1983), 882–888.

[17] Moon, Y. Intimate Exchanges: Using Computers to Elicit Self-Disclosure from Consumers. *Journal of Consumer Research 26*, 4 (2000), 323–339.

[18] Mori, M. Bukimi No Tani (The Uncanny Valley). *Energy 7*, 4 (1970), 33–35.

[19] Naimi, T.S., Brewer, R.D., Mokdad, A., Denny, C., Serdula, M.K., and Marks, J.S. Binge Drinking Among US Adults. *JAMA: The Journal of the American Medical Association 289*, 1 (2003), 70–75.

[20] Nass, C. and Moon, Y. Machines and Mindlessness: Social Responses to Computers. *Journal of Social Issues 56*, 1 (2000), 81–103.

[21] Nass, C., Steuer, J., and Tauber, E.R. Computers are Social Actors. *Proceedings of the ACM CHI Conference on Human Factors in Computing Systems*, ACM Press (1994), 72–78.

[22] Nederhof, A.J. Methods of Coping with Social Desirability Bias: A Review. *European Journal of Social Psychology 15*, (1985), 263–280.

[23] Nunamaker Jr, J.F., Derrick, D.C., Elkins, A.C., Burgoon, J.K., and Patton, M.W. Embodied Conversational Agent-Based Kiosk for Automated Interviewing. *Journal of Management Information Systems 28*, 1 (2011), 17–48.

[24] O'Brien, C. Facebook Messenger chief says platform's 34,000 chatbots are finally improving user experience.

VentureBeat, 2016. https://venturebeat.com/2016/11/11/facebook-messenger-chief-says-platforms-34000-chatbots-are-finally-improving-user-experience/.

[25] Pearce, W.B. and Sharp, S.M. Self-disclosing Communication. *Journal of Communication 23*, 4 (1973), 409–425.

[26] Perkins, H.W. and Berkowitz, A.D. Perceiving the Community Norms of Alcohol Use among Students: Some Research Implications for Campus Alcohol Education Programming*. *International Journal of the Addictions 21*, 9–10 (1986), 961–976.

[27] Pickard, M.D. Persuasive Embodied Agents: Using Embodied Agents to Change People's Behavior, Beliefs, and Assessments. 2012.

[28] Pickard, M.D., Burns, M.B., and Moffitt, K.C. A Theoretical Justification for Using Embodied Conversational Agents (ECAs) to Augment Accounting-Related Interviews. *Journal of Information Systems 27*, 2 (2013), 159–176.

[29] Pickard, M.D., Roster, C.A., and Chen, Y. Revealing Sensitive Information in Personal Interviews: Is Self-disclosure Easier with Humans or Avatars and under What Conditions? *Computers in Human Behavior 65*, (2016), 23–30.

[30] Posey, C., Lowry, P.B., Roberts, T.L., and Ellis, T.S. Proposing the Online Community Self-Disclosure Model. *European Journal of Information Systems 19*, 2 (2010), 181–195.

[31] Reeves, B. and Nass, C. *The Media Equation: How People Treat Computers, Television, and New Media Like Real People and Places*. Cambridge University Press, 2003.

[32] Richman, W.L., Kiesler, S., Weisband, S., and Drasgow, F. A meta-analytic study of social desirability distortion in computer-administered questionnaires, traditional questionnaires, and interviews. *Journal of Applied Psychology 84*, 5 (1999), 754–775.

[33] Sakshaug, J.W., Yan, T., and Tourangeau, R. Nonresponse Error, Measurement Error, And Mode Of Data Collection: Tradeoffs in a Multi-mode Survey of Sensitive and Non-sensitive Items. *Public Opinion Quarterly 74*, 5 (2011), 907–933.

[34] Short, J., Williams, E., and Christie, B. *The Social Psychology of Telecommunications*. Wiley, London, 1976.

[35] Sproull, L., Subramani, M., Kiesler, S., Walker, J., and Waters, K. When the Interface Is a Face. *Human-Computer Interaction 11*, 2 (1996), 97–124.

[36] Tinwell, A., Grimshaw, M., Nabi, D.A., and Williams, A. Facial Expression of Emotion and Perception of the Uncanny Valley in Virtual Characters. *Computers in Human Behavior 27*, 2 (2011), 741–749.

[37] Tourangeau, R., Couper, M.P., and Steiger, D.M. Humanizing Self-Administered Surveys: Experiments on Social Presence in Web and IVR surveys. *Computers in Human Behavior 19*, 1 (2003), 1–24.

[38] Tourangeau, R. and Smith, T.W. Asking Sensitive Questions: The Impact of Data Collection Mode, Question Format, and Question Context. *Public Opinion Quarterly 60*, 2 (1996), 275–304.

[39] Tourangeau, R. and Yan, T. Sensitive Questions in Surveys. *Psychological Bulletin 133*, 5 (2007), 859–883.

[40] Walker, J., Sproull, L., and Subramani, R. Using a human face in an interface. *Proceedings of the Conference on Human Factors in Computers*, (1994), 85–91.

[41] Wallace, R. The Anatomy of A.L.I.C.E. 2004.

[42] Wechsler, H., Lee, J.E., Kuo, M., Seibring, M., Nelson, T.F., and Lee, H. Trends in college binge drinking during a period of increased prevention efforts. Findings from 4 Harvard School of Public Health College Alcohol Study surveys: 1993-2001. *Journal of American college health : J of ACH 50*, 5 (2002), 203–217.

[43] Weizenbaum, J. ELIZA: A computer program for the study of natural language communication between man and machine. *Communications of the ACM 9*, 1 (1966), 36–45.

[44] Wilcox, B. ChatScript. 2017. https://github.com/bwilcox-1234/ChatScript.

[45] Yaghoubzadeh, R., Kramer, M., Pitsch, K., and Kopp, S. Virtual Agents as Daily Assistants for Elderly or Cognitively Impaired People: Studies on Acceptance and Interaction Feasibility. In R. Aylett, B. Krenn, C. Pelachaud and H. Shimodaira, eds., *Intelligent Virtual Agents*. Springer Berlin Heidelberg, Berlin Heidelberg, 2013, 79–91.