12-12-2022

Conversational Agents, Conversational Relevance, and Disclosure: Comparing the Effectiveness of Chatbots and SVITs in Eliciting Sensitive Information

Benjamin Richardson
Brigham Young University, benjum@byu.edu

Ryan Schuetzler
Brigham Young University, ryan.schuetzler@byu.edu

Conversational Agents, Conversational Relevance, and Disclosure: Comparing the Effectiveness of Chatbots and SVITs in Eliciting Sensitive Information

Benjamin Richardson
Brigham Young University
benjum@byu.edu

Ryan Schuetzler
Brigham Young University
ryan.schuetzler@byu.edu

ABSTRACT

Conversational agents (CAs) in various forms are used in a variety of information systems. An abundance of prior research has focused on evaluating the various traits that make CAs effective. Most studies assume, however, that increasing the anthropomorphism of an agent will improve its performance. In a sensitive information disclosure task, that may not always be the case. We leverage self-disclosure, social desirability, and social presence theories to predict how differing modes of conversational agents affect information disclosure. In this paper, we propose a laboratory experiment to compare how the mode of a given CA—text-based chatbot or voice-based smart speaker—paired with either high or low levels of conversational relevance, affects the disclosure of personally sensitive information. In addition to understanding influences on disclosure, we aim to break down the mechanisms through which CA design influences disclosure.

Keywords

Conversational agents, chatbots, smart speakers, disclosure, conversational relevance, social desirability

INTRODUCTION

The commonality of conversational agents (CAs) and their use in a variety of contexts has steadily grown as businesses and individuals adopt interfaces such as chatbots and smart voice-interaction technologies (SVITs) (Langevin, et al. 2021). Such communication technologies have become increasingly available, and as a result have replaced many in-person communications. Human-computer interactions in the form of conversational agents offer less costly, and potentially more effective means of obtaining and eliciting information from users (Pickard & Roster, 2020). The choice of interaction media can influence user perceptions, attitudes, and behaviors, and it is valuable to understand how the interaction changes by medium.

While prior research has thoroughly investigated the impact that specific designs decisions of CAs have on users (Abdul-Kader & Woods, 2015; Hussain, Ameri Sianaki, & Ababneh, 2019), very little research compares multiple types of CAs against each other in determining their ability to elicit sensitive information from users. This research aims to address this gap by empirically studying the differences in sensitive information elicitation by chatbots and smart voice-interaction technologies (SVITs). The following research question guides this work:

How does the mode of a CA and its level of conversational relevance influence disclosure of personally sensitive information?

BACKGROUND, LITERATURE REVIEW, AND HYPOTHESES

Chatbots and SVITs

Conversational agents (CAs) are systems that replicate human-to-human communication via text or oral speech using natural language processing, artificial intelligence, and machine learning (Khanna et al., 2015). Text-based CAs (or chatbots) enable interaction between users and machines through natural written language within integrated messaging systems (Rapp, Amon, Curti, & Boldi, 2021). SVITs are another kind of CA that serve as spoken dialogue systems often built into smartphones or smart speakers. They understand voice commands, respond through digital voices, and often handle tasks like monitoring home automated devices, calendars, email, and other information retrieval tasks (Adamopoulou, Eleni, & Moussiades, 2020).

Model of Disclosure for Conversational Agents

As these technologies develop, they are being used in increasingly novel and useful applications. One of these applications currently being researched is the use of chatbots and SVITs for interviewing (Pickard, Roster, & Chen, 2016; Schuetzler, Giboney, Grimes, & Nunamaker, 2018). If these technologies are to be used for interviewing and gathering potentially sensitive information, their ability to elicit accurate information will be crucial. For effective interviewing to take place, users must be willing to self-disclose. Self-disclosure can be defined as the extent to which users voluntarily and knowingly share information about themselves (Pearce & Sharp, 1973). However, challenges to disclosure may arise when the information being disclosed presents the user in a perceived negative light (Tourangeau & Smith, 1996). Therefore, the impact that various characteristics have on user disclosure is vital to consider. We propose a model

(Figure 1) that helps explain the variation in self-disclosure as manipulated by the mode of a CA and its level of conversational relevance.

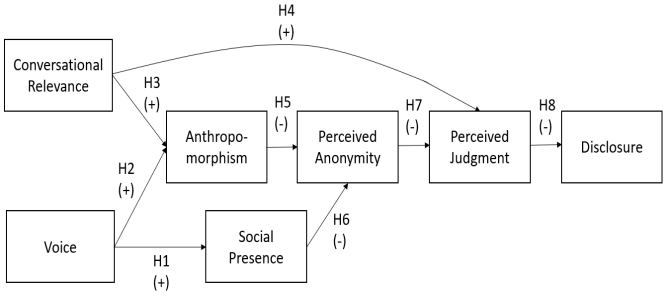

Figure 1. Model of Disclosure for Conversational Agents

Effect of Interview Modality on Anthropomorphism and Social Presence

When evaluating self-disclosure between different chatbot media, a few previous studies have explored questions that directly compare chatbots to SVITs. One study's results suggest that individuals interviewed using text-based mobile devices may provide more reliable and precise answers than those in more traditional spoken interviews (Schober, et al. 2015, p. 1). Additionally, voice-based agents have been shown to have a stronger anthropomorphic cue than chatbots because of their humanlike application of verbal speech (Widener & Lim, 2020). Based on the foundation of previous research, we hypothesize that SVITs will create a greater level of social presence and anthropomorphism than text-based chatbots:

H1. Social presence is higher for voice-based interactions than for text-based interactions.

H2. Anthropomorphism is higher for voice-based interactions than for text-based interactions.

Effect of Conversational Relevance on Anthropomorphism and Perceived Judgment

One key aspect of human-agent interactions is conversational relevance. Conversational relevance is strong when the response to a message or exhibition of conversational feedback is relevant and related to the current conversational topic (Giora, 1997). In CAs, perceptions of relevance are created by the agent's ability to respond contingently and appropriately to input from the user. When an agent can effectively communicate by holding a relevant conversation, users view it as being better able to understand them, thereby amplifying the sense of anthropomorphism (Pickard, Roster, & Chen, 2016). Non-relevant responses, on the other hand, provide the perception that the communication partner is disconnected from the user and more robotic. Such a disconnect diminishes the sense that the agent is human-like, and as a result directly influences the responses given by the interviewee.

We hypothesize that conversational relevance has a direct effect on both anthropomorphism and perceived judgment. When an agent has a high level of conversational relevance, users may perceive that the agent understands, internalizes, and has feelings toward the user's responses. This study proposes that as conversational relevance increases, the level of anthropomorphism also increases, effectively enhancing perceived judgment which diminishes self-disclosure:

H3. Increased conversational relevance increases the sense of anthropomorphism in the CA.

H4. Increased conversational relevance directly increases perceived judgment.

Effect of Anthropomorphism on Perceived Anonymity

Some literature suggests that perceiving the CA as possessing human-like characteristics—such as empathy and self-disclosure—improves the conversational experience, thereby encouraging the users to develop increased trust, tolerance, and openness toward the CA (Rapp, Curti, & Boldi, 2021). Despite the significant role that humanness plays in developing trust and enhancing the user experience, other research has highlighted that in contexts where eliciting potentially sensitive or embarrassing information is key, incorporating CAs that are less anthropomorphic may prove more effective (Schuetzler, Giboney, Grimes, Nunamaker, 2018). We propose that the mechanism for this difference is partially explained by the influence of anthropomorphism on *perceived anonymity*—the perception that one's identity is unknown. Perceiving the system as having more humanlike traits invites a Computers-Are-Social-Actors mindlessness in the interaction (Kim & Sundar, 2012), increasing also the unconscious perception that the agent can identify the user as an individual. The concept of anonymity relies on there being another entity to know one's identity. Identifying an individual, or de-anonymizing them, is one potential anthropomorphic trait that could be attributed to a more anthropomorphic chatbot:

H5. Higher anthropomorphism decreases user perceptions of anonymity.

Effect of Social Presence on Perceived Anonymity

Social presence is the sense of personal connection an individual feels with their communication partner (Short & Williams, 1976). Some research has shown that increased social presence may have a negative effect as individuals evaluate the social desirability of their responses and what ways their responses may affect their communication partner's opinion of them (Walker, Sproull, & Subramani, 1994; Schuetzler, Giboney, Grimes, & Nunamaker, 2018). This phenomenon, called social desirability, can be described as the way in which individuals prefer to be seen by others (Crowne & Marlowe, 1960) and often refers to the systematic tendency to give overly positive answers that frame the respondent in a positive light (Paulhus, 2002). In the case of high social presence, users will often adjust their responses to match what they perceive a socially desirable response to be. We anticipate that the

closer the user feels to the conversation partner, the less anonymous the user will feel (Wagenknecht, Teubner, & Weinhardt, 2016). Therefore, we expect that high levels of social presence will have a direct effect on perceived anonymity:

H6. Increased social presence has a negative effect on perceived anonymity.

Effect of Perceived Anonymity on Perceived Judgment

Perceived judgment is the perception that one is being evaluated or judged by another entity. This judgment usually has a negative connotation. In interpersonal interactions, perceived judgment has important implications for decision making and behavior. Feeling judged by one's doctor, for example, influences the likelihood that a patient will attempt weight loss (though not their success) (Gudzune, Bennett, & Cooper, 2014).

Perceived anonymity has been well documented to increase self-disclosure (e.g., Joinson, 2001; Rains, 2014). We propose that the mechanism for this effect is mediated through perceived judgment. Prior research has found that when an individual feels their identity is anonymous, they may communicate more boldly than they would in a face-to-face situation—a phenomenon called the online disinhibition effect (Clark-Gordon, Bowman, Goodboy, & Wright, 2019). Thus, when an individual feels they are anonymous in their communication, they may feel they cannot be judged at an individual level by the person (or chatbot) they are communicating with.

H7. An increase in perceived anonymity reduces perceived judgment.

Effect of Perceived Judgment on Information Disclosure

The goal of our study is to understand the mechanism by which design features of a conversational agent influence sensitive information disclosure. Prior research has shown a relationship (Schuetzler et al., 2018). Pickard et al. (2016) showed in a survey that participants would feel more comfortable disclosing to a CA compared to a human. In their study, 51% of participants expressed that a lack of judgment from a CA is one major reason for this preference. Lind et al. refer to this judgment as the evaluative capability of the agent, and that more humanlike systems are seen as having greater evaluative capability, and therefore reduce disclosure of sensitive information. (Lind, Schober, Conrad, Reichert, 2013)

H8. As perceived judgment increases, self-disclosure is diminished.

METHOD

Study Design

To evaluate our hypotheses, participants will be randomly assigned into one of four experimental conditions: an interview with a text-based chatbot or an interview with an SVIT, with each mode of communication having either high or low levels of conversational relevance (see *Table 1*). The sample will be drawn from business school students at a large U.S.-based private university with a strict honor code. A post-experiment survey will measure the mediating constructs.

Conversational Agent	Low Conversational Relevance	High Conversational Relevance
Chatbot (text)	n = 35	n = 35
SVIT (voice)	n = 35	n = 35

Table 1. Breakdown of Conditions

Identifying Sensitive Questions

To effectively assess the proposed hypotheses, it is imperative that significantly sensitive topics be addressed in the interviews. Based on the academic context of the student population, this study will evaluate the extent of academic dishonesty, or cheating, that individual university students participate in. Additionally, this study will be performed at a university with a strict honor code which largely emphasizes academic integrity. We anticipate this topic will prove to be significantly sensitive for the student population.

Conversational Agent Development

The conversational agents will be developed using the Microsoft Bot Framework. The voice interaction agents will be deployed to Amazon Echo smart speakers.

Procedure

The randomized condition will be assigned following participant registration for the study. Each participant who reports for their assigned experiment time will participate in either an online interview with a chatbot or an interview with an SVIT. The same interview questions will be asked in a sequential order for each of the conditions. Following completion of the interview, participants will be directed to complete a post-experiment survey.

Interview Questions

There are a total of 16 interview questions derived from a 2012 study survey that assessed the frequency and scope of cheating of 4,316 students from 10 high schools (Galloway, 2012). These questions, most of which are of sensitive nature to actively enrolled students, allow us to measure how willing the study participants are to disclose potentially embarrassing information.

Post-Experiment Survey

Following completion of the interview, study participants will be directed to the post-experiment survey. The survey consists of validated measures that allow us to gather basic demographic information and evaluate our proposed

model. This series of questions will enable us to establish individual baselines for socially desirable responding.

ANALYSIS

Following the completion of the experiment, the data from the responses across the various treatments will be standardized for each interview question. Due to the unknown extent of academic dishonesty among the population, and the exorbitant lengths that obtaining such information would require, it will be assumed that due to the random assignment of conditions to participants, the mean values for the sensitive topics will not be systematically different across the various conditions. Therefore, any significant differences comparing the conditions can be attributed to the study's manipulations. By using the collective trends among the samples, we can estimate the impact of the condition manipulations on overall disclosure.

CONCLUSION

CAs in their various forms are being increasingly leveraged to gather sensitive information, and as such, it is imperative to consider the effects that different modes of CAs, as well as their level of social presence, have on self-disclosure. Based on prior literature, it is evident that in contexts such as customer service, entertainment, and data retrieval, CAs with a greater level of anthropomorphism may better facilitate the user experience. However, this study aims to investigate contexts where eliciting honest, yet socially delicate information is critical (e.g., medical interviews, customs processing, criminal investigations, etc.).

Should the proposed hypotheses prove veritable, this study will provide a significant contribution to the field of human-robot interaction, further crediting social presence theory and presenting more effective use cases for text and voice-based CAs. We anticipate that further research in this vein may be needed to further understand the implications that varying levels of conversational relevance, social presence, and anonymity may have on CAs and their ability to elicit sensitive information, especially in contexts that differ from the one presented in this study.

REFERENCES

1. Abdul-Kader, S. A., and Woods, J. C. (2015) "Survey on Chatbot Design Techniques in Speech Conversation Systems," *International Journal of Advanced Computer Science and Applications* (6:7), The Science and Information (SAI) Organization.
2. Adamopoulou, E., and Moussiades, L. (2020) "Chatbots: History, Technology, and Applications," *Machine Learning with Applications* (2), p. 100006.
3. Asgeirsdottir, R. L., Vésteinsdóttir, V., and Thorsdottir, F. (2016) "Short Form Development of the Balanced Inventory of Desirable Responding: Applying Confirmatory Factor Analysis, Item Response Theory, and Cognitive Interviews to Scale Reduction," *Personality and Individual Differences* (96), pp. 212–221.
4. "Automated Assistance in E-Commerce: An Approach Based on Category-Sensitive Retrieval | SpringerLink." (n.d.).
5. "Avatar E-Mail versus Traditional e-Mail: Perceptual Difference and Media Selection Difference - ScienceDirect." (n.d.).
6. Clark-Gordon, C. V., Bowman, N. D., Goodboy, A. K., and Wright, A. 2019. "Anonymity and Online Self-Disclosure: A Meta-Analysis," *Communication Reports* (32:2), Routledge, pp. 98–111.
7. "Conversational Agents in Healthcare: A Systematic Review | Journal of the American Medical Informatics Association | Oxford Academic." (n.d.). (https://academic.oup.com/jamia/article/25/9/1248/5052181?login=false, accessed April 28, 2022).
8. Crowne, D. P., and Marlowe, D. (1960) "A New Scale of Social Desirability Independent of Psychopathology," *Journal of Consulting Psychology* (24:4), US: American Psychological Association, pp. 349–354.
9. Foehr, J., and Germelmann, C. C. (2020) "Alexa, Can I Trust You? Exploring Consumer Paths to Trust in Smart Voice-Interaction Technologies," *Journal of the Association for Consumer Research* (5:2), pp. 181–205.
10. Galloway, M. K. (2012) "Cheating in Advantaged High Schools: Prevalence, Justifications, and Possibilities for Change," *Ethics & Behavior* (22:5), Routledge, pp. 378–399.
11. Giora, R. (1997) "Discourse Coherence and Theory of Relevance: Stumbling Blocks in Search of a Unified Theory," *Journal of Pragmatics* (27:1), pp. 17–34.
12. Go, E., and Sundar, S. S. (2019) "Humanizing Chatbots: The Effects of Visual, Identity and Conversational Cues on Humanness Perceptions," *Computers in Human Behavior* (97), pp. 304–316.
13. Gratch, J., Wang, N., Gerten, J., Fast, E., & Duffy, R. (2007) "Creating rapport with virtual agents," In *International workshop on intelligent virtual agents* (pp. 125-138). Springer, Berlin, Heidelberg.
14. Gudzune, K. A., Bennett, W. L., Cooper, L. A., & Bleich, S. N. (2014). "Perceived judgment about weight can negatively influence weight loss: A cross-sectional study of overweight and obese patients." Preventive Medicine, 62, 103–107.
15. Hussain, S., Ameri Sianaki, O., and Ababneh, N. (2019) "A Survey on Conversational Agents/Chatbots Classification and Design Techniques," in *Web, Artificial Intelligence and Network Applications*, Advances in Intelligent Systems and Computing, L. Barolli, M. Takizawa, F. Xhafa, and T. Enokido (eds.),

Cham: Springer International Publishing, pp. 946–956.

16. Joinson, A. N. (2001). "Self-disclosure in computer-mediated communication: The role of self-awareness and visual anonymity." European Journal of Social Psychology, 31(2), 177–192.
17. Khanna, A., Pandey, B., Vashishta, K., Kalia, K., Bhale, P., and Das, T. (2015) "A Study of Today's A.I. through Chatbots and Rediscovery of Machine Intelligence," *International Journal of U- and e-Service, Science and Technology* (8), pp. 277–284.
18. Kim, Y., & Sundar, S. S. (2012). Anthropomorphism of computers: Is it mindful or mindless? *Computers in Human Behavior*, *28*(1), 241–250.
19. Langevin, R., Lordon, R. J., Avrahami, T., Cowan, B. R., Hirsch, T., and Hsieh, G. (2021) "Heuristic Evaluation of Conversational Agents," in *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, CHI '21, New York, NY, USA: Association for Computing Machinery, May 6, pp. 1–15.
20. Li, X., Hess, T. J., and Valacich, J. S. (2008) "Why Do We Trust New Technology? A Study of Initial Trust Formation with Organizational Information Systems," *The Journal of Strategic Information Systems* (17:1), pp. 39–71.
21. Lind, L. H., Schober, M. F., Conrad, F. G., & Reichert, H. (2013). Why Do Survey Respondents Disclose More When Computers Ask the Questions? Public Opinion Quarterly, 77(4), 888–935.
22. Linnemann, G. A., and Jucks, R. (2018) "'Can I Trust the Spoken Dialogue System Because It Uses the Same Words as I Do?'—Influence of Lexically Aligned Spoken Dialogue Systems on Trustworthiness and User Satisfaction," *Interacting with Computers* (30:3), pp. 173–186.
23. "Logic and Conversation in: Speech Acts." (n.d.). (https://brill.com/view/book/edcoll/9789004368811/BP000003.xml, accessed April 28, 2022).
24. Marlow, D., and Crowne, D. P. (1961) "Social Desirability and Response to Perceived Situational Demands," *Journal of Consulting Psychology* (25:2), US: American Psychological Association, pp. 109–115.
25. Paulhus, D. L. (2002) "Socially desirable responding: The evolution of a construct." In H. Braun, D. N. Jackson, & D. E. Wiley (Eds.), The role of constructs in psychological and educational measurement (pp. 67-88). Hillsdale, NJ: Erlbaum
26. Paulhus, D. L. (1986) "Self-Deception and Impression Management in Test Responses," in *Personality Assessment via Questionnaires: Current Issues in Theory and Measurement*, A. Angleitner and J. S. Wiggins (eds.), Berlin, Heidelberg: Springer, pp. 143–165.
27. Pickard, M. D., and Roster, C. A. (2020) "Using Computer Automated Systems to Conduct Personal Interviews: Does the Mere Presence of a Human Face Inhibit Disclosure?," *Computers in Human Behavior* (105), p. 106197.
28. Pickard, M. D., Roster, C. A., and Chen, Y. (2016) "Revealing Sensitive Information in Personal Interviews: Is Self-Disclosure Easier with Humans or Avatars and under What Conditions?," *Computers in Human Behavior* (65), pp. 23–30.
29. Rains, S. A. (2014). "The Implications of Stigma and Anonymity for Self-Disclosure in Health Blogs," Health Communication, 29(1), 23–31.
30. Rapp, A., Curti, L., and Boldi, A. (2021) "The Human Side of Human-Chatbot Interaction: A Systematic Literature Review of Ten Years of Research on Text-Based Chatbots," *International Journal of Human-Computer Studies* (151), p. 102630.
31. Schober, M. F., Conrad, F. G., Antoun, C., Ehlen, P., Fail, S., Hupp, A. L., Johnston, M., Vickers, L., Yan, H. Y., and Zhang, C. (2015) "Precision and Disclosure in Text and Voice Interviews on Smartphones," *PLOS ONE* (10:6), Public Library of Science, p. e0128337.
32. Schuetzler, R. M., Giboney, J. S., Grimes, G. M., and Nunamaker, J. F. (2018) "The Influence of Conversational Agent Embodiment and Conversational Relevance on Socially Desirable Responding," *Decision Support Systems* (114), pp. 94–102.
33. J. Short, E. Williams, B. Christie, (1976) The Social Psychology of Telecommunications, Wiley, London,
34. Tourangeau, R., and Smith, T. W. (1996) "Asking Sensitive Questions: The Impact of Data Collection Mode, Question Format, and Question of Context," *Public Opinion Quarterly* (60:2), pp. 275–304.
35. Wagenknecht, T., Teubner, T., and Weinhardt, C. 2016. "The Impact of Anonymity on Communication Persuasiveness in Online Participation," *ICIS 2016 Proceedings*.
36. Walkerl, J. H., Sproull, L., and Subramani, R. (n.d.). *Using a Human Face in an Interface*, p. 7.
37. Widener, A., and Lim, S. (2020) "Need to Belong, Privacy Concerns and Self-Disclosure in AI Chatbot Interaction," *Journal of Digital Contents Society* (21:12), pp. 2203–2210