Prestige Questions, Online Agents, and Gender-Driven Differences in Disclosure

Abstract. This work considers the possibility of using virtual agents to encourage disclosure for sensitive information. In particular, this research used "prestige questions", which asked participants to disclose information relevant to their socioeconomic status, such as financial credit, as well as university attendance, and mortgage or rent payments they could afford. We explored the potential for agents to enhance disclosure compared to conventional web-forms, due to their ability to incorporate minimal levels of rapport. To consider this possibility, agents were framed as artificially intelligent versus avatars controlled by a real human, and we compared these conditions to a version of the financial questionnaire with no agent. Additionally, we examine the differences in disclosure between men and women in these conditions. Analyses reveled that agents (either AI- or human-framed) evoked greater disclosure compared to the no agent condition. However, there was some evidence that human-framed agents evoked greater lying. Thus, users in general responded more socially to the presence of a human- or AI-framed agent, and the benefits and costs of this approach were made apparent. The results are discussed in terms of rapport and anonymity.

Keywords: Virtual Agents; Human-Agent Experimentation; Disclosure.

1 Introduction

Any automated system that aims to understand or interact with humans needs to have a good model of how those humans will interact with it. Critically, when such a system's function is to collect data, it is imperative that the data it collects are accurate and readily reported. This is especially important in medical fields, where information may be critical in correctly diagnosing patients. Indeed, failure to provide fully honest responses in medical interviews can result in serious consequences for patient health. Therefore, much research has considered how to gain more detailed and honest medical histories, especially sensitive information, from patients [1].

It is also helpful in other, less life-threatening domains—specifically, disclosure of honest financial data allows automated systems to provide helpful information to users to aid them in financial planning and other tasks. However, these kinds of information are often difficult to gather due to social effects that may encourage people to misrepresent or refuse to answer questions about their health or finances [2]. These so-called "prestige questions" are ones that many users may be reluctant to answer, or, when they do, to answer truthfully.

Generally, work has found that rapport and anonymity both foster more honest disclosure on sensitive topics. Honest disclosure requires both 1) willingness to answer sensitive questions, and, if willing to report on such sensitive questions, it then requires 2) answering those questions honestly. However, little to no research has explicitly disentangled honest disclosure of sensitive information in this way: by examining willingness to simply answer sensitive questions *separately* from willingness to answer such questions honestly. Furthermore, most research has focused on reporting of stigmatized health related information, leaving the issue of reporting sensitive information related to socioeconomic status unexplored.

To address these issues, we presented a series of prestige questions, where we asked participants to answer questions about their financial status. After a series of required questions, participants were asked if they were willing or unwilling to answer more sensitive questions that would directly reveal their socioeconomic status. Later, they were asked if they were honest on the required questions. This allowed us to disentangle honest disclosure of sensitive information as well as address disclosure in the domain of socioeconomic status.

The value of automated agents to influence these issues should not be overlooked, given the ubiquitous nature of online interactions. Indeed, the positive effect of agents on the generation of trust when discussing finance has been demonstrated [3]. Previous work [4] provides early models for predicting disclosure based on trust of various commercial websites and aims to model different users. However, this work does not condition on the basis of a virtual human partner, something explicitly shown to affect disclosure in other contexts [5]. In this context, we consider the effect of rapport (even minimal rapport) by examining the presence of an agent (vs. a standard web form). We consider the effect of anonymity by comparing AI-framed agents to human-framed agents. This work thus expands the body of knowledge on human behavior in conjunction with virtual partners into the financial domain.

2   Related Work

2.1   Disclosure, Rapport, and Anonymity

Financial disclosure is a relatively untested domain, and even less work has been done therein that attempts to utilize virtual human partners to effect desired outcomes. Although the use of computer agents to increase disclosure has been examined, previous studies has focused on intimate self-disclosure, rather than on financial or prestige questions [6]. And yet, automated "tele-operator" systems have been in use for years, and their importance has been well established [7]. Indeed, these systems are often seen as critical, complicated, and often protected by various patents [8].

Yet, the effects of *how* these systems influence user behavior, and if there are design principles that improve outcome measures (such as disclosure), are still relatively untested. The decision on whether to provide live phone support, automated tele-

phone systems, chat-bots, or more realistic agents is an important one. Since these decisions are likely to affect the level of rapport generated between an agent/service provider and the user, care must be taken to understand the effects on user behavior [9].

Current evidence suggests that different types of interaction produce different levels of disclosure by affecting two key psychological factors: rapport and anonymity – or the sense that one's identity is protected. Generally, research has shown that greater feelings of rapport lead people to disclose more [5, 10, 11]. Indeed, because computer- and self-administered assessment lack any human element, these traditional assessments do not evoke the same feelings of rapport or social connection. Specifically, when there is not a human or human-like agent present in some way, shape, or form, people feel less socially-connected during the assessment [9, 12, 13].

Besides rapport, anonymity is another psychological factor that leads to differences in disclosure. Indeed, much of the research exploring the effect of anonymity on disclosure has done so by contrasting different assessment methods such as computer-mediated interviews, face-to-face interviews, computer-administered assessment, and standard self-assessment. Research shows that computer-mediated interviews are felt to be more anonymous than face-to-face interviews, just as computer-administered assessment is when compared to self-administered assessment by paper-and-pencil, and this resultant anonymity leads to increased disclosure [14].

2.2 Gender-Based Differences in Behavior

It is important to realize that, although individual differences in user experience are always to be expected, there may be systemic differences based on user gender (and often, based on gender match/mismatch with the perceived gender of the agent). Features such as social distance also often play a role in changing the rapport and behavior of users with others [15]. Often, these system design decisions can make huge differences in the usability and effectiveness of that system [16]. Specifically, the gender of a partner can have substantial effects in social situations, especially ones in which prestige questions are likely to be involved [17].

These predictions are quite distinct between men and women. Evidence suggests that having an opposite-gendered tutor (among heterosexual participants) may increase rapport behavior in certain instructional interactions. This was found to be especially true for female users interacting with a male instructor [18]. Furthermore, in conversations when financial questions were involved—and thus earning potential was discussed—women were far more concerned than men in considering their partners [19]. Expectedly, this can lead to increased deception on the part of males when presenting this information, especially in situations where that information is often shared (such as online dating) [20]. In essence, these works lead us to the idea that if our goal is to increase disclosure of financial information, it pays to be cognizant of the gender-based differences that may arise.

3   Experimental & System Design

To approach these questions of disclosure, we designed a 3-condition study to determine the effects of a virtual assistant when users were asked various financial questions. In all conditions, users were merely told that they would be asked some questions about their finances as part of a study to determine how best to provide information for financial planning. This survey was designed using the Qualtrics survey software. The survey involved a series of questions that were determined to be prestige questions due to the social nature of the queries.

The questions were divided into 5 sections, of which users saw a subset based on their prior answers. The sections were: 1) House ownership, 2) Apartment rental, 3) Past education, 4) Current education, and 5) Credit history. At the start of several sections, users were asked questions that determined which of the subsequent sections they would view. For example, before starting section one, they were asked if they lived in a home, apartment, or dorm room. Users that answered "home" would see section 1 but not section 2. Users that answered "apartment" would see section 2 but not section 1. And users that answered "dorm room" would see neither of those sections. Similar questions gated sections 3 and 4 (Are you attending college/have you previously attended college?) and section 5 (Do you currently have any credit cards?).

Each of these sections contained a number of questions designed to be somewhat uncomfortable but anonymous. For example, the house section asked about current mortgage payments, their amounts and interest rates, and the total value of the home. Near the end of each section, there was also a question that was designed to break the anonymity. These questions were considered the disclosure questions for the purposes of this study. In the house section, this question was "Will you give us your address so we can check the value of the house using Zillow (an online service) to verify this? (If yes, we'll ask you for that information later)." Note that regardless of their answer, this data was *never* collected, and we did not violate the anonymity of the participants. A summary of the sections and questions is listed in Table 1.

Table 1. Survey Disclosure Questions

Section	Disclosure Question
Home ownership	Address of home?
Apartment rental	Address of apartment?
Past school	School name?
Current school	School name?
Credit history	Approval to run a credit check?

At the conclusion of these 5 sections, all participants were debriefed and asked another set of questions regarding whether or not they had previously lied on any of

the previous questions (as in [4]), as well as some basic demographic information, including their gender.

Our experimental manipulation took place during the 5 primary sections. In the control condition, there was nothing added to the survey, and it was simply a regular online form. In the "Agent" and "Human" conditions, the survey was augmented with a simple virtual agent that remained at the top of the survey. This agent can be seen in Figure 1. The left image depicts the agent in the "Agent" condition, wherein the photo was made to look like a drawing. The right image depicts the agent in the "Human" condition, wherein the photo remained lifelike (both photos were simple stock photos freely available without copyright).

Fig. 1. "Agent" framing condition with automated remarks (left), and "Human" framing condition with additional remarks after a choice has been made (right)

The behavior of the agent in the "Agent" and "Human" conditions was identical. The agent regularly sent messages in the chat window, providing additional explanation about the questions. They also reacted to the user's actions, providing context-specific information based on the user's choices. An example of this can be seen in Figure 2, which takes place after the user has clicked the "Home" option. This behavior was accomplished using the Qualtrics JavaScript API.

The only differences between the two experimental conditions were, (as stated previously) the pictures, and an initial framing description that users saw at the beginning of the survey. In the "Agent" condition, users were told that they were about to interact with an "artificial intelligence (AI) agent", and in the "Human" condition, they were told that they would interact with a "live support representative". Participants experienced a brief "connection delay" in both conditions as they were "matched" with their agent or representative. All participants were debriefed regarding the deception following the study.

4 Results

We recruited 381 participants (241 male, 140 female) using Amazon's Mechanical Turk (MTurk) service as subjects for our study. Their average age was 35.13 (SD =

10.40). Participants were recruited only from the United States (verified using IP address), and were required to have a 98% or higher approval rating using MTurk's reporting system.

Participants completed a manipulation check; specifically, they were asked to identify whether they completed the answers with a human, an agent, or just completed the form themselves. From the original 381 participants, 52 failed the manipulation check. These participants were excluded, leaving 329 for analysis.

4.1 Willingness to answer sensitive questions

A one-way ANOVA revealed a significant difference between conditions ($F_{(2, 326)} = 3.38$, $p = .04$; see Table 2). Follow-up contrasts indicate that, out of the 4 possible questions, participants were willing to provide significantly more answers to either a human ($t_{(326)} = 2.05$, $p = .04$) or an agent ($t_{(326)} = 2.10$, $p = .04$) than a form. However, human and virtual agents did not elicit different levels of willing to provide these answers ($t_{(326)} = 0.22$, $p = .82$). In a 3 (Condition: Form, Agent, or Human) x 2 (Gender: Male or Female) ANOVA to test gender, there was no main effect or interaction with gender (Fs < 1.47, ps > .23).

Table 2. Proportion of Answers Disclosed

Condition	Form (Control)	Agent Framing	Human Framing
Percentage of possible questions answered	47.5%	55%	55%

To investigate whether there was an effect of gender on any of the 4 individual questions, we ran log-linear analyses on participants' willingness to answer each of the 4 questions separately. These analyses revealed that women's responses drove the above effects more than men's for 3 of the 4 questions, but men's responses drove the effect more than women's for the final question. First considering participants' willingness to allow the system to do a credit check, the analysis revealed a significant interaction with gender ($G2(7) = 14.78$, $p = .04$). Indeed, women tended to consent to credit checks more with an agent (14%) or human (13%) than on a form (6%; $\chi^2(2) = 1.72$, $p = .42$); however, men tended to show the opposite: they consented to credit checks on the form (15%) more than with an agent (6%) or human (7%; $\chi^2(2) = 4.29$, $p = .12$).

Similarly, the effect of condition on participants' willingness to provide the name of their former college or university also depended on gender ($G2(7) = 17.40$, $p = .02$). Women tended to be more willing to share the name of their former school with an agent (35%) than a human (29%) or on a form (31%; $\chi^2(2) = 1.38$, $p = .50$); however, men were significantly more willing to provide the school name on the form (54%) than to an agent (34%) or human (29%; $\chi^2(2) = 6.08$, $p = .048$). There was also a trend for a condition by gender interaction on participants' willingness to provide

the name of the current school (G2(7) = 10.12, p = .18). Again, women tended to be more willing to provide the name of their current school to a human (43%) or agent (39%) than on a form (12%; $\chi^2(2)$ = 2.10, p = .35), whereas men tended to show the reverse: they were less willing to provide the name to a human (12%) than on the form (44%) or to an agent (38%; $\chi^2(2)$ = 2.80, p = .25).

In contrast, while the effect of condition on participants' willingness to provide their address was also significantly qualified by gender (G2(7) = 16.84, p = .02), the effect this time was more driven by men. Men were significantly more willing to provide their address to a human (90%) or agent (72%) than on a form (62%; $\chi^2(2)$ = 9.24, p = .01), whereas women were equally willing across conditions (74% vs 72% vs 70%, respectively; $\chi^2(2)$ = 0.21, p = .90).

4.2   Reported lying on questions

Additionally, chi-squared tests were run to test the effect of condition on reported lying on each item. Participants were significantly more likely to lie about their credit limit when asked by a human (8%) then when asked by an agent (1%) or on a form (2%; $\chi^2(2)$ = 6.93, p = .03). Again, this effect was again driven by female participants more than males. There was a marginally significant interaction with gender (G2(7) = 12.44, p = .087), whereby women were marginally more likely to say they lied to a human (9%) then when asked by an agent (0%) or on a form (2%; $\chi^2(2)$ = 5.60, p = .06), and men just showed a trend to lie more to a human (6%) than an agent (2%) or on a form (2%; $\chi^2(2)$ = 1.79, p = .41). The effect of condition failed to reach statistical significance for lying on the remaining items (χ^2s < 2.88, p > .23).

5   Discussion

These results suggest not only that there is a substantial benefit to providing an agent or human-controlled avatar to increase user experience and disclosure, but also that gender-based differences are very significant. These disclosure results are summarized in Table 3.

Table 3. Disclosure Results Summary

Question	Overall, more likely to provide answer to:	Males only:	Females only:
Total number of answers	Agents and humans*	Agents and humans	Agents and humans
Credit Check	Agents and humans	Web form	Agents and humans
Past school name	Agents and humans*	Web form*	Agents and humans
Current school name	Agents and humans	Web form	Agents and humans
Address	Agents and humans*	Agents and humans*	Equally likely

*Significant at p < 0.05

The core result of these findings is that the use of a virtual human partner increases disclosure in the domain of financial questions. This result, already shown with agent-assisted disclosure in the medical domain, is another, stronger statement on the usefulness of virtual agent assistants. Generating rapport with a virtual agent—even a relatively simple one like the agent used in this study—will yield actionable results when attempting to get answers to prestige questions that may be hard to get otherwise.

But this result is also tempered by demographic effects—increasing the social presence of a task has both benefits and drawbacks. The agent/avatar used in this study was clearly female, and its gender may have had a large effect on the perceptions and actions of the users according to their own genders—manipulating the gender of the agent is a good target for future work. As previous research has suggested, prestige questions that reveal information about financial security are often very important for females looking for a long-term partner. Therefore, it should not be surprising that the males in our study were reluctant to reveal the answers to questions like their credit history and school history to the female agent/avatar. However, the males in this study were willing to reveal their addresses to the agent/avatar. While this result is somewhat unexpected, it is possible that the address is seen as a "safer" option—whereas the very mention of a credit score or school name invites immediate social judgment, determining social status using merely an address does require additional steps (someone would have to search the address using an online service, judge the neighborhood, etc.). Therefore, this question may be perceived as less potentially damaging to males looking to impress.

Indeed, the results on lying do seem in line with this hypothesis. It was the question on credit, not on housing addresses, that prompted lying behavior. When speaking to a human that could potentially judge the users for their perceived poor credit, users lied more than they did to either an agent or a web form. Since none of the other questions prompted (self-reported) lying behavior to a large degree, one possible conclusion is that the credit limit question is seen as potentially the most socially damaging. This is in line with previous research on loss of anonymity.

But, regardless of the cause of the lying behavior, it should be concluded that there are serious gender-driven effects when asking prestige questions that should be taken into account. While it is still true that the use of an automated agent/avatar clearly increased disclosure overall in this study, that effect may backfire when used on males. In future work, androgynous or male agents could be used to further determine if female agents drive this behavior, or if gender match/mismatch does.

In sum, the addition of a very simple automated agent was able to produce dramatic changes in the behavior of users on an online form. Prestige questions, much as other questions that people may be reluctant to answer, are often a critical point in data collection. This method paves to the way to collect more data, more accurately, and to target different demographics with data-driven, intelligent techniques.

References

1. Maguire, Peter, et al. "Helping cancer patients disclose their concerns." European Journal of Cancer 32.1 (1996): 78-81.
2. Gabler, Neal. "The Secret Shame of Middle-Class Americans." The Atlantic (May 2016).
3. Bickmore, Timothy, and Justine Cassell. "Relational agents: a model and implementation of building user trust." Proceedings of the SIGCHI conference on Human factors in computing systems. ACM, 2001.
4. Metzger, Miriam J. "Privacy, trust, and disclosure: Exploring barriers to electronic commerce." Journal of Computer-Mediated Communication 9.4 (2004): 00-00.
5. Lucas, Gale M., et al. "It's only a computer: virtual humans increase willingness to disclose." Computers in Human Behavior 37 (2014): 94-100.
6. Moon, Youngme. "Intimate exchanges: Using computers to elicit self-disclosure from consumers." Journal of consumer research 26.4 (2000): 323-339.
7. Bose, Ranjit. "Customer relationship management: key components for IT success." Industrial management & Data systems 102.2 (2002): 89-97.
8. Fawcett, Philip E., and Christopher Blomfield-Brown. "System and method for providing automated customer support." U.S. Patent No. 5,678,002. 14 Oct. 1997.
9. Gratch, Jonathan, et al. "Creating rapport with virtual agents." International Workshop on Intelligent Virtual Agents. Springer Berlin Heidelberg, 2007.
10. Burgoon, Judee K., Laura K. Guerrero, and Kory Floyd. Nonverbal communication. Routledge, 2016.
11. Hall, Judith A., Jinni A. Harrigan, and Robert Rosenthal. "Nonverbal behavior in clinician—patient interaction." Applied and Preventive Psychology 4.1 (1995): 21-37.
12. DeVault, David, et al. "SimSensei Kiosk: A virtual human interviewer for healthcare decision support." Proceedings of the 2014 international conference on Autonomous agents and multi-agent systems. International Foundation for Autonomous Agents and Multiagent Systems, 2014.
13. Gratch, Jonathan, et al. "Creating rapport with virtual agents." International Workshop on Intelligent Virtual Agents. Springer Berlin Heidelberg, 2007.
14. Weisband, Suzanne, and Sara Kiesler. "Self disclosure on computer forms: Meta-analysis and implications." Proceedings of the SIGCHI conference on human factors in computing systems. ACM, 1996.
15. Hoffman, Elizabeth, Kevin McCabe, and Vernon L. Smith. "Social distance and other-regarding behavior in dictator games." The American Economic Review 86.3 (1996): 653-660.
16. Howell, Mark, Steve Love, and Mark Turner. "User characteristics and performance with automated mobile phone systems." International Journal of Mobile Communications 6.1 (2008): 1-15.
17. Evers, Catharine, et al. "Anger and social appraisal: a" spicy" sex difference?." Emotion 5.3 (2005): 258.
18. Krämer, Nicole C., et al. "Closing the gender gap in STEM with friendly male instructors? On the effects of rapport behavior and gender of a virtual agent in an instructional interaction." Computers & Education 99 (2016): 1-13.
19. Buss, David M., and David P. Schmitt. "Sexual strategies theory: an evolutionary perspective on human mating." Psychological review 100.2 (1993): 204.
20. Toma, Catalina L., Jeffrey T. Hancock, and Nicole B. Ellison. "Separating fact from fiction: An examination of deceptive self-presentation in online dating profiles." Personality and Social Psychology Bulletin 34.8 (2008): 1023-1036.