It's Only a Computer: The Impact of Human-agent Interaction in Clinical Interviews

Jonathan Gratch
Institute for Creative Technologies
12015 Waterfront Dr., Los Angeles
gratch@ict.usc.edu

Gale Lucas
Institute for Creative Technologies
12015 Waterfront Dr., Los Angeles
glucas@usc.edu

Aisha King
Bard College
30 Campus Rd, Annandale, NY
ak4628@bard.edu

Louis-Philippe Morency
Institute for Creative Technologies
12015 Waterfront Dr., Los Angeles
morency@ict.usc.edu

ABSTRACT
Research has begun to explore the use of virtual humans (VHs) in medical interviews [1]. When designed as supportive and "safe" interaction partners, VHs may improve such screenings by encouraging patients to disclose more personal information [2-3]. In medical contexts, patients often feel resistance to self-disclosure and engage in impression management to be viewed more positively by healthcare providers. This paper provides the first empirical evidence that VHs can reduce such resistance and impression management. In the context of health-screening interviews, we report a study in which participants interacted with a VH that was either teleo-operated by humans (Wizard-of-Oz) or fully-automated (AI). Independently, we manipulated whether participants believed the VH was controlled by humans or automation. As predicted, participants who believed they were interacting with a computer reported lower resistance to self-disclosure, lower impression management and higher system usability than those who believed they were interacting with a human operator. Whether the virtual human was actually operated by a human or AI only affected ratings of the system's usability. These results suggest that automated VHs can help overcome a significant barrier to obtaining truthful patient information in medical domains.

Categories and Subject Descriptors
I.2.11 [**Artificial Intelligence**]: Human-agent Interaction – *virtual humans*

Keywords: Virtual humans, computer-assisted assessment, medical interviews, clinical interviews, self-disclosure, impression management

1. INTRODUCTION
Virtual humans are becoming an important tool for training across a number of contexts from military and business to medical and counseling. Users can learn a number of skills sets, from interpersonal skills to specific tasks, by interacting with virtual humans. As tools, virtual humans (VHs) could have some advantages over their "live" human counterparts. For example, using VHs in place of real people is usually less labor-intensive. Users' experiences can be better standardized with VHs than with human beings. VHs can also provide a "safe" environment to explore social interactions. This sense of safety can encourage learning or honest disclosure of important information.

The healthcare field, in particular, may benefit from the latter potential advantage of virtual humans: honest disclosure of personal information. 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 [4-5]. For example, physicians have been encouraged to establish *rapport,* or the experience of harmony, fluidity, synchrony, and flow felt during a conversation, with their patients [2]. Indeed, rapport fosters honesty [4-6].

Although a number of factors may contribute to patients providing more honest, detailed information to healthcare providers [6], psychological barriers to honest responding are primary factors that can be modified. At least two psychological factors can hinder honest responding in medical contexts: 1) *fear of self-disclosure*, and 2) *impression management*.

First, patients might not reveal personal information out of a fear of self-disclosure. They often hold back information because they feel afraid that they are being viewed negatively by the healthcare professional [7]. Patients may be particularly afraid to disclose personal, sensitive or stigmatizing information. Unfortunately, such information can be the most important for them to disclose to healthcare professionals.

Second, patients engage in impression management [8], only disclosing information that will lead healthcare providers to view them positively. They try to selectively represent themselves and their behaviors in ways that they believe will make healthcare professionals view them positively.

Through fear of self-disclosure and impression management, healthcare professionals may get a very inaccurate picture of the patient and his or her medical history, which can have serious health consequences. We argue that VHs could be used to reduce these psychological barriers to honest responding. Specifically, we contend that interacting with a VH will result in lower fear of self-disclosure and reduced impression management.

"The possibility that people would tell an impartial machine personal or embarrassing things about themselves, without fear of negative evaluation, has been raised since the first uses of computers for communication" [9]. Since these initial uses of

Appears in: *Alessio Lomuscio, Paul Scerri, Ana Bazzan, and Michael Huhns (eds.), Proceedings of the 13th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2014), May 5-9, 2014, Paris, France.*

computers for assessment, researchers have established a robust literature investigating the benefits of using computer-administered assessment methods to solicit more honest, detailed responding. A meta-analysis of this literature found that computer-administered assessment methods lead to greater disclosure of personal information than traditional non-computerized methods [9].

The effect of computer-administered assessment on honest responding is particularly strong in medical and mental health contexts due to the intimate nature of the information required. Consider, for example, the task that healthcare professionals have when assessing suicide risk. As suicide is highly stigmatized in most cultures, patients are often afraid to admit to such thoughts in face-to-face interviews. However, when asked to disclose information about suicidal thoughts by computer-administered assessment methods, participants not only felt more positively about the assessment than with traditional assessment methods, but, more importantly, they gave more honest answers [10]. Although computer-administered assessments can improve honest responding for even mundane private information [11-12], these effects are especially strong when the information is illegal, unethical, or culturally stigmatized [9]. As many behaviors that harm mental and physical health fall into this category (e.g., drug use or unsafe sex), computer-administered assessments can be especially important in health domains.

Accompanying their meta-analysis [9], Weisband and Kiesler offer an explanation for why people give more honest responses to a computer: when compared to traditional methods, computer-administered assessment formats allow for a "sense of invulnerability to criticism, an illusion of privacy, the impression that responses 'disappear' into the computer". Several studies have confirmed that respondents perceive computer-administered assessments to be more anonymous than traditional paper-and-pencil assessments, and computer-mediated interviews to be more anonymous than face-to-face interviews [11-15].

Computer-administered assessments can make participants feel more anonymous in a number of ways, and any of these aspects may increase honest responding. Specifically, computer-administered assessments remove two factors that have been shown to undermine anonymity and honest responding: 1) the "mere presence" of another person, and 2) the "mere belief" that responses are being observed and judged.

First, merely being present in the same room with another person ("mere presence") can reduce anonymity and thereby decrease honest responding. For example, when participants interacted with each other via a computer-mediated system (i.e., in different rooms), they felt more anonymous and disclosed more, compared to when participants interacted with each other face-to-face [12]. In other words, simply sitting across from another human being can make people respond less honestly.

Second, the "mere belief" that another is – in the moment – evaluating their responses can discourage people from responding honestly. For example, people respond more honestly in face-to-face interviews if the interviewer remains unaware of their answers. Turner and colleagues employed a survey technique where, after the interviewer asks a question, the respondent answers on a separate sheet of paper that is then sealed in an envelope [16] (the approach is especially useful when respondents have poor reading skills). Even though the person knows their answers will eventually be seen and judged, removing the "mere belief" that that their responses are judged in the moment can encourage people to respond more honestly.

Whereas Turner et al.'s survey technique [16] removed only "mere belief", other survey formats allow for removal of both "mere presence" and "mere belief". Indeed, self-administered questionnaires, for example, are completed without an interviewer (hence they are "self"-administered), and participants' responses are not observed while they complete the assessment. Accordingly, compared to face-to-face interviews, self-administered questionnaires increase honest responding across a wide range of domains [17-19]. Although it is unclear which aspect of anonymity (removal of "mere presence" or "mere belief") is driving this effect of self-administered questionnaires on honesty, the effect does seem to be broad. Indeed, self-administered questionnaires can reduce each of the aforementioned barriers to honest responding: fear of self-disclosure and impression management. First, self-administered assessments can help patients overcome their fear of disclosing stigmatized personal information, which they could have hidden, such as reporting that they had an abortion [17]. Second, during such assessments, participants do not slant their reports of health-related habits as much, reporting greater alcohol and drug use as well as a higher number of sexual partners [18-19].

Across self-administered, computer-administered, and even face-to-face assessments, it seems that the "mere belief" that another human will observe patients' responses makes them less willing to be honest in what they report. Moreover, removing this "mere belief," along with the "mere presence" of another human, helps to make both computer- and self-administered assessments seem "safe" and anonymous. Accordingly, participants are more likely to respond honestly, overcoming fear of negative evaluation and desire to engage in impression management.

2. THE PRESENT RESEARCH

Even though computer- and self-administered assessments increase honest responding, it is possible that other formats could result in even more forthcoming disclosures. Computer- and self-administered assessments allow for removal of the "mere presence" of another human, but – in these formats – the experience of rapport is removed right along with "mere presence". Because they lack any human element, traditional computer- and self- administered assessments do not evoke any feelings of social connection. This lack of rapport cuts against the aforementioned benefits of anonymity in healthcare contexts. While the anonymity of computer-administered assessments increases honest responding of sensitive health information, decrements in felt connection decrease self-disclosure [4-6]. Therefore, computer-administered assessments may not evoke as much honest responding as would be possible if these assessments encouraged feelings of rapport.

However, with the advent of virtual humans (VHs), computer-administered assessments can now evoke feelings of rapport without the "mere presence" of another human being. Indeed, when VHs have been designed to be human-like in their relational behaviors, they evoke feelings of rapport. Compared to agents without relational behaviors, relational agents have been found to increase such feelings, and in turn, liking of the agent [1-3]. Moreover, because patients can interact with VHs without another human present, VHs could evoke a feeling of rapport *without* "mere presence" of another human. Without "mere presence," patients could also feel more anonymous than they do during face-to-face interviews. Therefore, when relational VHs are used to give computer-administered assessments, respondents could experience a sense of rapport without sacrificing the sense of anonymity that comes with computer-administered assessments.

Specifically, as a form of computer-administered assessment, interviewing with a VH should be experienced as anonymous in both senses: removal of "mere presence" and "mere belief". When interacting with a VH, participants usually believe (or can be led to believe) that their responses are not currently being observed by another human being. Additionally, participants who are interacting with a VH are not placed face-to-face with a real person. Yet, like interacting with a real human, interacting with a relational VH would also create feelings of social connection. Experiencing rapport in the absence of "mere presence" and "mere belief" could be optimal for eliciting honesty. Therefore, in efforts towards more open responding in medical contexts, VHs could provide the best of computer-assisted and face-to-face assessment.

Given that VHs could evoke a sense of anonymity (absence of "mere presence" and "mere belief") like other computer-administered assessments while at the same time preserving feelings of rapport that occur in human-human interactions, interviews with VHs should lower fear of self-disclosure compared to interacting with a human. If participants believe that no other human is observing their responses, they also feel that it is "safe" to express themselves without engaging in impression management. Likewise, because there is no one present to impress, absence of "mere presence" should also result in less impression management. Therefore, overall, interviews with VHs should result in decreased fear of self-disclosure and impression management compared to interviews with a human.

However, it is possible that any one of these three factors –rapport, "mere presence" or "mere belief" – could affect honesty in responding. For example, removal of "mere belief" alone could reduce fear of self-disclosure and impression management. If this is the case, interviews with VHs should reduce such fear and impression management not only compared to interviews with humans, as we argue above, but also compared to interviews with a VH that participants believe is teleo-operated by a human. That is, if removal of "mere belief" is sufficient to elicit greater honesty, interacting with a VH should result in more honest responding when participants believe the VH is automated than when they believe it is controlled by a human being.

We argue that removal of "mere belief" is sufficient. Even if human were – in fact – to observe participants' responses during interviews with a VH, if participants *believe* that there is no human present, they should respond more honestly. For instance, if a VH interviewer were to be teleo-operating (like a puppet - or like the Wizard-of-Oz), but the participants believed it was operated by a computer system, the assumed anonymity afforded by this belief should be sufficient to lower fear of self-disclosure and impression management. In this paper, we use such a paradigm to establish that believing that one is interacting with an automated VH can result in more honest responding. Therefore, this paper provides the first empirical test of the possibility that, because they remove the "mere belief" that responses are being judged, VHs can reduce fear of self-disclosure and impression management.

Specifically, to isolate the impact of "mere belief" on such honest responding, other factors that could affect honesty – rapport and "mere presence" – need to be held constant. Therefore, we employed a VH interviewer that was designed to be relational and lack "mere presence" (no real human was present during the interview), and for which "mere belief" could be manipulated. Participants were told that the VH was either operated by a computer or a human. Through this manipulation of frame, we were able to test whether believing that responses are not currently being observed by another human being (computer frame) helps reduce fear of self-disclosure and impression management compared to believing that responses were being observed (human frame).

Given that *belief* is assumed to be essential to "mere *belief*", only beliefs about whether the VH is operated by computer or human should matter. Whether the VH is actually operated by the computer or teleo-operated by a human should be irrelevant for fear of self-disclosure and impression management. Therefore, in addition to frame, we also manipulated whether participants actually interacted with a VH that was teleo-operated by human (Wizard-of-Oz) or fully-automated (AI). We hypothesized that people who believed they were interacting with a computer during an interview with a VH (computer frame) would exhibit less fear of self-disclosure and impression management than those who believed that they were interacting with a human during the interview with the VH (human frame). In contrast to this expected effect of "mere belief", we predict that whether participants interact with a VH that is either actually teleo-operated by humans (Wizard-of-Oz) or, in fact, fully-automated (AI) would have no effect on fear of self-disclosure and impression management.

To test these hypotheses, we conducted a study in which participants interacted with a VH during a semi-structured interview. The VH was either teleo-operated by humans (Wizard of Oz) or fully-automated (AI). Independently, we manipulated whether participants believed the VH was controlled by humans (human frame) or a computer program (computer frame). After the interview, participants completed measures of fear of self-disclosure [20] and impression management [21]. Additionally, we investigated whether participants would also rate the system itself more highly when it was framed as controlled by a computer; accordingly, participants were finally asked to complete the System Usability Scale [22].

3. METHOD

3.1 Design

Two hundred and thirty nine participants (149 males, 90 females) were recruited via Craigslist. All participants who met requirements (i.e. age between 18 and 65, and adequate eyesight) were accepted.

In this study, all participants interacted with a VH. Participants were randomly assigned to a computer frame or a human frame: some participants were *told* that they would be interacting with an automated system, and others were *told* that they would be interacting with a human-controlled system. The experimenter introduced participants to the system using graphic representations as well as verbal descriptions that manipulated frame. Participants in both frames were told that there would not be any real human in the room with them, and thus their responses would be "anonymous" (in that "mere presence" was removed). Although anonymity in terms of "mere presence" was held constant across conditions, our frame manipulation varied in terms of "mere belief". In the computer frame, participants were led to believe that their responses *were not* being observed. They were told:

> …Our virtual human uses artificial intelligence to have a conversation with you, yet preserves your anonymity. The system gets audio and visual input from you. It uses a Speech Recognition tool to understand what you're

> saying, then uses a complex series of equations to choose the best way to respond...

In the human frame condition, participants were led to believe that their responses *were* being observed. They were told:

> … Our virtual human is like a puppet. It allows a person in another room have a conversation with you, yet preserves your anonymity… My colleague will be sitting in the other room, and be able to see and hear you on this screen. She has access to a set of pre-recorded questions and responses that will be used to have a conversation with you....

In reality, some participants interacted with our VH when it was controlled by a human (Wizard-of-Oz method), whereas other participants interacted with our VH when her responses were fully automated (AI method). All participants were randomly assigned to either the computer or human frame, regardless of whether they experienced the WoZ or AI method. Accordingly, the experiment consisted of a 2 (frame: computer vs. human) × 2 (method: WoZ vs. AI) design. In all interactions, the VH interacts verbally and nonverbally in a semi-structured manner with a participant.

In WoZ, Wizards were blind to framing condition. However, in 21 WoZ sessions, Wizards learned the condition; these were excluded. Additionally, after interacting with our VH, participants completed a manipulation check: they selected whether the VH was "controlled by a computer" or "controlled by a human". Those who failed the manipulation check were also excluded. Together, these exclusions left 154 subjects, 77 in each condition. This includes 57 Wizard of Oz interactions and 97 Artificial Intelligence interactions. In 9 interactions with the Artificial Intelligence, the system failed; these sessions were also excluded.

3.2 The Virtual Interviewer: SimSensei

Across all conditions of the experiment, participants engaged in a spoken interaction with the SimeSensei relational VH (Figure 1). System details are described elsewhere (see [23]). Human interviewers employ a variety of skills to reduce patient fear of self-disclosure, and several research efforts have examined how to automate and incorporate these skills into VH systems. For example, the rapport agent by Gratch and colleagues [2] uses vision and prosodic analysis to provide active listening behaviors (e.g., smiles, head nods and postural mimicry). The Relational Agents work of Bickmore and colleagues combines these nonverbal skills with verbal techniques including expressions of empathy, social dialogue and reciprocal self-disclosure [1].

Figure 1. SimSensei, the Virtual Human Interviewer

Empirically, these techniques have been shown to increase self-disclosure, as well as feelings of rapport, self-efficacy and trust [1-3].

SimSensei builds on this existing technological base and is part of a larger research effort to examine the feasibility of virtual health agents for mental health screening. This kiosk-based system is aimed at clients resistant to seeking traditional care and combines the advantages of traditional web-based self-administered screening, which allows for anonymity, with a VH interface, which fosters the rapport-building techniques employed by the Rapport- and Relational-agent work.

The virtual human conducts a semi-structured screening interview with a user via spoken language. The interview is structured around a series of agent-initiated questions organized into phases: initially there is a rapport-building phase where the agent asks general introductory questions (e.g., "Where are you from originally?); this is followed by a clinical phase where the agent asks a series of questions about symptoms (e.g., "How easy is it for you to get a good night's sleep?"); finally, the agent ends with a series of more positive questions designed to return the patient to a more positive mood (e.g., "What are you most proud of?"). At each phase, the agent can ask follow-up questions (e.g., "Can you tell me about that?"), provide verbal empathetic feedback (e.g., "I'm sorry to hear that"), and produce nonverbal behaviors (e.g., nods, expressions) for active and empathetic listening.

The system itself was developed in an iterative fashion: an analysis of face-to-face interviews informed the design of a semi-automated system, which in turn, informed the creation of a fully-automated AI prototype. Analysis of the face-to-face interviews produced a library of questions, follow-up questions and verbal and nonverbal feedback behaviors that could be used by the agent. The semi-automated system used human operators (Wizard-of-Oz method), operating the system from another room, drive the selection and timing of these library-elements during a series of interviews. While the basic order of questions was structured into phases, wizards had to decide 1) when to take the turn from the user, 2) if to ask a follow-up question, 3) if to provide verbal empathetic feedback and 4) when to provide nonverbal behaviors such as nods and smiles. The low-level realization of these behaviors was determined by automation. These wizard choices serve as a corpus of training data from which a fully-automated system was created. Both the semi-automated (Wizard-of-Oz) and fully-automated (AI) versions of the interviewer are used in this study. As mentioned, Wizards were blind to frame condition. For additional information about development and implementation of this system, see DeVault et al. [23].

3.3 Measures and Procedure

After giving consent and receiving an explanation of the system, which served as a manipulation of frame, participants completed demographic questionnaires. Then participants engaged in an interview with the virtual human. After the dialogue concluded, participants completed a number of questionnaires. First, to measure fear of self-disclosure, participants completed a modified version of the Brief Fear of Negative Evaluation Scale – Straightforward Items [20]. Specifically, items were modified to refer to their fear that participants felt during the interaction with our VH. For example, on a scale from 1 (strongly disagree) to 5 (strongly agree), participants responded to items like "I was afraid the interviewer noticing my shortcomings." Participants then completed the Balanced Inventory of Desirable Responding

(BIDR), measuring impression management and self-deception [21]. On a scale from 1 (strongly disagree) to 5 (strongly agree), participants responded to items like "I never regret my decisions" and "I am a completely rational person," respectively. Although we were only interested in the effect on impression management, for completeness, we also report analyses for self-deception. Finally, participants completed the System Usability Scale [22] with respect to our virtual human dialogue system. Participants rated the system on a scale from 1 (strongly disagree) to 5 (strongly agree) with items such as, "I thought the system was easy to use," for example. Upon completion of the questionnaires, participants were debriefed, thanked and paid.

4. RESULTS

We used Fear of Negative Evaluation (FNE) to measure fear of self-disclosure, Impression Management (IM) subscale to index impression management, and System Usability (SU) scale to system usability. For completeness, Self-deception (SD) subscale was also analyzed. For each of these scales, participants' total scores were calculated for analysis.[1] Table 1 displays group means for total scores and associated standard errors for each measure.

Table 1. Total Score Means and Standard Errors for Frame Conditions and Method Conditions

	Frame		Method	
	Computer	Human	WoZ	AI
FNE	16.54 (.69)	18.61 (.63)*	17.19 (.76)	17.95 (.53)
IM	56.29 (1.33)	59.76 (1.16)*	57.68 (1.39)	58.38 (1.08)
SD	66.72 (1.24)	68.59 (1.08)	69.03 (1.30)	66.29 (1.01)
SU	72.46 (1.64)	68.54 (1.44)†	73.85 (1.72)	67.16 (1.34)*

Note. † = $p \leq .075$. * = $p \leq .05$.

First, we conducted a 2 (frame: computer vs. human) × 2 (method: WoZ vs. AI) ANOVA on participants' total fear of negative evaluation (FNE) scores. There was only a significant main effect of frame, $F(1, 130) = 4.90$, $p = .03$, $d = 0.75$, such that those who received the computer frame reported lower fear of self-disclosure than those with the human frame (see Figure 2).[2]

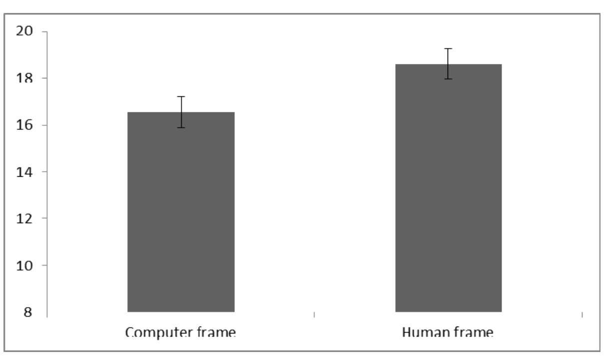

Figure 2. Fear of negative evaluation scores by frame.

Results of an ANOVA on participants' total impression management (IM) scores revealed a main effect of frame, $F(1, 141) = 3.88$, $p = .05$, $d = 0.37$, such that participants who received the computer frame reported lower impression management after the interaction than those who received the human frame (see Figure 3). No other effects approached significance, and a compulsory parallel ANOVA on total self-deception (SD) scores yielded no significant effects, $F\text{s} < 2.77$, $p\text{s} > .10$.

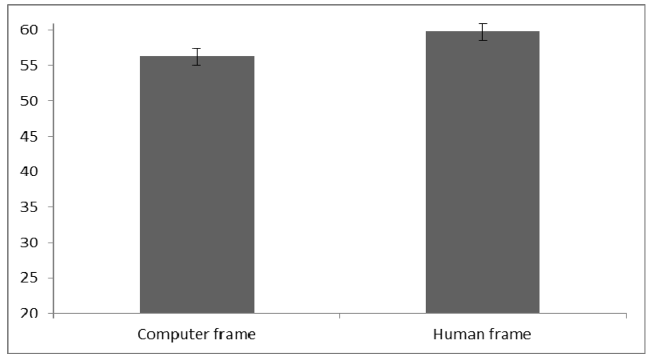

Figure 3. Impression management scores by frame.

Finally, analysis of system usability (SU) scores yielded a marginal main effect of frame, $F(1, 141) = 3.23$, $p = .075$, $d = 0.30$, such that participants who received the computer frame rated the virtual human dialogue system as higher in usability than those who received the human frame (see Figure 4).

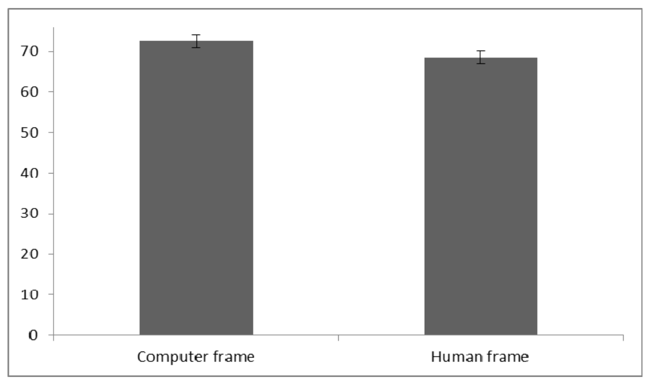

Figure 4. System Usability scores by frame.

This analysis also yielded a main effect of method, $F(1, 141) = 9.42$, $p = .003$, $d = 0.52$, those who experienced WoZ rated the system as more usable than those in the AI group (see Figure 5).[3]

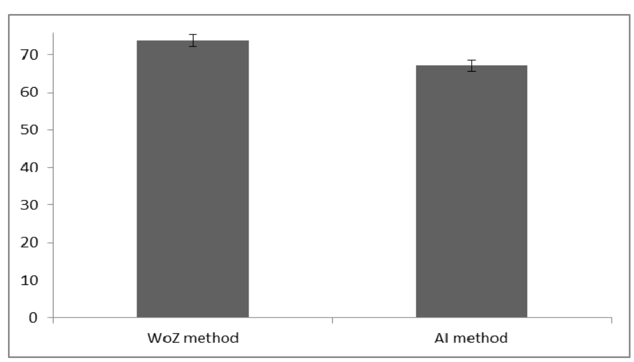

Figure 5. System Usability scores by method.

5. DISCUSSION

Results confirmed our hypotheses and showed that *automated* virtual humans (VHs) are able to reduce fear of self-disclosure and impression management. Participants who were told that the VH interviewer was fully automated (computer frame) reported significantly lower fear of negative evaluation and impression management than participants who were told that the VH interviewer was teleo-operated by humans (human frame). This provides clear evidence that VHs can retain the benefits of computer-administered assessments (i.e., removal of both the "mere presence" of an interviewer and the "mere belief" that one is being judged) while simultaneously incorporating the benefits of relational skills (e.g., rapport and social dialog).

In addition to manipulating frame, the present study also manipulated whether participants actually interacted with a VH that was teleo-operated by human (Wizard-of-Oz) or fully-automated (AI). We assumed that only the *belief* that responses will not be observed would affect honest responding, not whether the responses would actually be observed or not. As expected, there was no effect of or interaction with method (Wizard-of-Oz versus AI) on fear of self-disclosure or impression management.

While frame condition did have this effect on impression management, the effect on self-deception did not approach significance. As we merely analyzed this subscale of BIDR for completeness, we did not expect an effect of "mere belief" on self-deception. Indeed, feeling anonymous should encourage honest reporting to others, not necessarily honesty with oneself. This divergence between the impression management and self-deception subscales of the Balanced Inventory of Desirable Responding does, however, add further validity to the distinction made between these two sub-factors of desirable responding [21].

We also explored the effect of believing that the VH interviewer was fully automated versus teleo-operated by humans on ratings of the VH system itself. Participants rated the system marginally more positively when they believed it was fully automated than when they believed it was teleo-operated. This marginally significant effect may have been due to a halo effect, such that participants extended (or misattributed) their positive feelings of "safety" and anonymity to the system itself. Alternatively, it may simply be that the system seems more impressive when one believes that it is automated due to the novelty of such automation.

In contrast to the effect of "mere belief" on system usability, the actual method (Wizard-of-Oz versus AI) employed to operate the VH had an opposite effect on these ratings of the system. Regardless of frame, Participants who interacted with a VH interviewer that was fully automated rated the system as lower in usability than those who interacted with the VH when it was teleo-operated by humans. As the automated system is still in development, the automated VH made more errors compared to when the VH was teleo-operated by humans. It is likely that this higher prevalence of errors is sufficient to explain the lower ratings of system usability. Importantly, however, this greater incidence of errors committed by the automated VH did not result in any greater fear of self-disclosure or impression management. This bodes well for use of VHs -even those still in development- for eliciting personal information.

Indeed, this paper provides the first empirical evidence that VHs can reduce fear of self-disclosure and impression management in a clinical interview context. Additionally, we show that VHs are able to have this impact because they allow patients to feel as though their responses are not currently being judged. This belief that responses were not being judged allowed participants to feel less fearful of disclosing information and reduced desire to make a good impression. Participants unsolicited anecdotal remarks echo the importance of "mere belief" that this study establishes:

> "I wish you hadn't told me that other people were in the other room listening in. It was weird, like, I don't even know these people. I would have said a lot more stuff if they weren't there."

> "[I] would have felt more comfortable if nobody was watching."

More specifically, the present results demonstrate that "mere belief" must be removed in order to elicit the fullest level of disclosure. That is, even when "mere presence" is removed and rapport is felt, removing the "mere belief" that responses are being judged evokes more honest responses. We were able to establish this by manipulating "mere belief" while holding "mere presence" and rapport constant. We held "mere presence" constant by not allowing other humans in the room during the interview in either frame condition. Likewise, the same relational VH was used to elicit equivalent feelings of rapport across both frame conditions. Therefore, the only difference between frames was the "mere belief" that another human was observing responses during the interview session.

Although this work establishes the importance of "mere belief" for honest responding, it only compares "mere belief" to no "mere belief" in the case that "mere presence" has been removed and rapport has been established. It does not test "mere belief" with any other combination of "mere presence" and rapport, nor does it test any other combinations of these three factors. For example, to test "mere presence", participants could interview with a VH while someone else was in the room or not. However, participants in both conditions of such study would have to be told the VH interviewer was teleo-operated by a human. Otherwise, the factor of "mere belief" would be manipulated along with – and thus confounded with – this manipulation of "mere presence". Overall, research should consider different combinations of these three factors that can break down psychological barriers to honest responding.

Additionally, considering these factors of "mere presence", "mere belief" and rapport can also help to illuminate differences between assessment formats: 1) face-to-face interviews, 2) computer-mediated interviews, 3) self-administered assessments, 4) computer-administered self-assessments, and 5) VH-administered interviews. Indeed, each of these formats seem to have different benefits and drawbacks. First, the primary advantage that face-to-face interviews have is rapport. Although participants would feel greater rapport in face-to-face interviews than during computer-administered self-assessments and traditional self-assessments, face-to-face interviews trade-off anonymity for this rapport [11-15]. Another key advantage of face-to-face interviews is when respondents have poor literacy, and thus unable to properly understand self-assessments [16], although they similarly trade-off anonymity for this benefit. Conversely, although such self-administered assessments provide anonymity, they would not evoke feelings of rapport like face-to-face or VH-administered interviews.

Traditional self-administered questionnaires also have another serious drawback: only relatively simple questionnaires can be given in this format. Computer-assisted self-assessments overcome this problem, making it possible to use very complex questionnaires without the aid of an interviewer. Additionally, although computer-mediated interviews (i.e., in different rooms) would allow for such complex questionnaires, these interviews remove only “mere presence” but not “mere belief”. Face-to-face interviews also allow for such complex questionnaires, but remove neither “mere presence” or “mere belief”.

However, VH-administered interviews could possibly allow for all of these benefits. VH interviewers allow for rapport and administration of more complex questionnaires without “mere presence” or “mere belief”. Therefore, additional research should consider how VH-administered interviews stack up against these more-established methods. Although research has found that rapport helps to elicit self-disclosure by comparing relational VH agents to non-relational agents [1-3], these types of VH have not been directly compared to traditional computer- and self-administered assessments. Such comparisons would investigate the role the *degree* of experienced rapport plays in how these assessments encourage honest responding.

VH-administered assessments may have their own drawbacks as well. Although, in our research, VHs who were framed as automated lowered impression management, VHs in general may evoke more impression management than self-administered assessments. Because computers can be viewed as if they are human, they can make people behave as though they are interacting with a real person [24]. Therefore, people may engage in impression management when a VH is present [25-26]. Future research should address this and investigate the circumstances that make VH interviewers more or less likely to evoke impression management. We begin this effort here by demonstrating that framing the VH as a computer rather than a human can reduce impression management.

Indeed, the power of VH interviewers to elicit more honest responding comes from the sense that no one is observing or judging. Not only does our effect of “mere belief” demonstrate this, but participants evidenced this with additional unsolicited anecdotal remarks. Specifically, those who believed the VH was automated stated:

> "This is way better than talking to a person. I don't really feel comfortable talking about personal stuff to other people."
>
> "A human being would be judgmental. I shared a lot of personal things, and it was because of that.”
>
> "It was helpful to have someone listen to me non-judgmentally."

The present work highlights the value of VH for clinical interviews. However, we have only participants’ reports of their inclination to disclose rather than more objective measures of self-disclosure. While it is valuable to demonstrate that people feel more comfortable disclosing, it would be useful to have objective indicators of disclosure as well. Additionally, given we relied solely on self-report measures, one might argue that differential self-report biases across conditions could account for our effects. However, given that BIDR’s measure of impression management is a “covert measure,” measuring the construct indirectly, such biases cannot account for the effect on impression management.

Another limitation of this work is that the scope of the experiment was restricted by practical constraints. For logistical reasons, participants fear of self-disclosure and impression management was only measured immediately after their interaction with our VH. However, some recent work suggests that the beneficial impact of interacting with a VH may not last over time. Vardoulakis found that social desirability biases (e.g., impression management) increased over time when patients interacting with a VH over time; in contrast, those interacting with a text-based system showed decreased social desirability over time [27]. Likewise, we limited our research to consider fear of self-disclosure and impression management. Constructs like trust and privacy would also be important to encourage honest responding; however, we did not collect data on these issues.

In this paper, we consider the effect of using VHs during a clinical interview that is psychological in nature. It is important to investigate the impact of using VHs in medical interviews that focus on physical health phenomena. VHs could be useful in a number of medical assessment domains. For example, VH interviewers may elicit more honest responses from cancer patients. Cancer patients’ honesty is required to properly adjust their treatment; however, these patients could also have especially heightened fears of disclosing this information to healthcare providers. A patient, for instance, might be afraid to disclose her concerns about her side effects because she is worried the cancer might progress if her oncologist, in turn, decides to lower her dose of radiation. However, information about the side effects she is experiencing could be important for her healthcare provider to know. Interviewing with a VH, instead, could ease patients’ fears of self-disclosure.

Successful VH-administered assessments could spawn additional uses of VHs in medicine. For example, VH patients could be used to role-play with healthcare professionals to improve their interviewing skills. Perhaps a specially-designed VH patient could even help to train medical interviewers to elicit greater honesty from patients. Just as VH-administered assessments can reduce patients’ fear of negative evaluation when providing personal information, using VH for this kind of role-playing during training exercises could also reduce healthcare professionals’ fear of being evaluated negatively by their peers [28].

The “possibility that people would tell an impartial machine personal or embarrassing things about themselves, *without fear of negative evaluation*” has been born out. Here we demonstrate that VHs can help overcome psychological barriers to honesty in clinical interviews. Providing more honest responses in medical interviews can help patients to receive better care and avoid serious health consequences. Therefore, the benefits to patients of VH-administrated clinical interviews could be quite substantial.

6. REFERENCES

[1] Bickmore, T., Gruber, A., & Picard, R. W. (2005). Establishing the computer-patient working alliance in automated health behavior change interventions. *Patient Education and Counseling*, 59(1), 21-30.

[2] Gratch, J., Wang, N., Gerten, J., & Fast, E. (2007). Creating rapport with virtual agents. Paper presented at the *7th International Conference on Intelligent Virtual Agents*, Paris, France.

[3] Gratch, J., Kang, S.-H., & Wang, N. (2013). Using social agents to explore theories of rapport and emotional resonance In J. Gratch & S. Marsella (Eds.), *Emotion in nature and artifact*: Oxford University Press.

[4] Maguire, P., Falkkner, A., Booth, K., Elliott, C., & Hillier, V. (1996). Helping cancer patients disclose their concern. *European Journal of Cancer, 32A*, 78-81.

[5] Roter, D, & Hall J. (1987). Physicians' interviewing styles and medical information obtained from patients. *Journal of General Internal Medicine, 2,* 325-329.

[6] Beckman, H., & Frankel, R. (1984). The effect of physician behavior on the collection of data. *Annals of Internal Medicine, 101*, 692-696.

[7] Farber, B. A. (2006). *Self-disclosure in Psychotherapy*. Guilford, New York, NY.

[8] Leary, M. R. & Kowalski, R. M. (1990). Impression management: Literature review and two-component model. *Psychological Bulletin, 107*, 34–47.

[9] Weisband, S., & Kiesler, S. (1996). Self-disclosure on computer forms: Meta analysis and implications. *CHI, 96,* 3-10.

[10] Greist, J. H., et al. (1973). A computer interview for suicide-risk prediction. *American Journal of Psychiatry, 130,* 1327-1332.

[11] Beckenbach, A. (*1995)*. Computer assisted questioning: The new survey methods in the perception of the respondent. *Bulletin de Méthodologie Sociologique, 48,* 82-100.

[12] Joinson, A.N. (2001). Self-disclosure in computer-mediated communication: The role of self-awareness and visual anonymity. external website. *European Journal of Social Psychology, 31,* 177-192.

[13] Baker, R.P. (1992). New Technology in Survey Research: Computer-Assisted Personal Interviewing (CAPI). *Social Science Computer Review, 10*, 145-157.

[14] Sebestik, J., Zelon, H., DeWitt, D., O'Reilly J.M., & McGowan, K. (1988). Initial Experiences with CAPI. *Proceedings of the Fourth Annual Research Conference,* March, Arlington, Virginia (pp.357- 371). Washington D.C.: U.S. Department of Commerce, Bureau of the Census.

[15] Thornberry, O., Rowe, B., & Biggar, R. (1990). Use of CAPI with the U.S. National Health Interview Survey. Paper presented at the World Congress of Sociology, Madrid.

[16] Turner, C.F., Lessler, J.T., & Devore, J.W. (1992). Effects of mode of administration and wording on reporting of drug use. In C.F. Turner, J.T. Lessler, & J.C. Gfroerer (Eds.), *Survey measurement of drug use: Methodological studies* (DHHS Publication No. ADM 92-1929, pp. 177-220). Rockville, MD: National Institute on Drug Abuse.

[17] London, K.A., & Williams, L.B. (1990). A Comparison of Abortion Underreporting in an In-Person Interview and a Self-Administered Questionnaire. Paper presented at the annual meeting of the Population Association of America, Toronto, Canada.

[18] Aquilino, W. S., & LoSciuto, L. A. (1990). Interview mode effects in drug use surveys. *Public Opinion Quarterly, 54,* 362-395.

[19] Tourangeau, R., & Smith, T. W. (1996). Asking sensitive questions: The impact of data collection mode, question format, and question context. *Public Opinion Quarterly, 60,* 275–304.

[20] Carleton, R. N., Collimore, K. C., McCabe, R. E., & Antony, M. M. (2011). Addressing revisions to the Brief Fear of Negative Evaluation scale: Measuring fear of negative evaluation across anxiety and mood disorders. *Journal of Anxiety Disorders, 25,* 822-828.

[21] Li, A., & Bagger, J. (2007). The balanced inventory of desirable responding: A reliability and generalization study. *Educational and Psychological Measurement, 40*, 131–141.

[22] Brooke, J. (1996). SUS: A "quick and dirty" usability scale. Jordan, P. W., Thomas, B., Weerdmeester, B. A., McClelland (eds.) *Usability Evaluation in Industry*, pp. 189-194. Taylor & Francis, London, UK.

[23] DeVault, D., et al. (*this volume*). SimInterviewer: A Virtual human interviewer for health decision support. *Proceedings of AAMAS'13*, Paris, France.

[24] Nass, C., Moon, Y., Fogg, B. J., Reeves, B., & Dryer, D. C. (1995). Can computer personality be human personalities? *International Journal of Human-Computer Studies, 43,* 223-239.

[25] Bailenson, J. N., Yee, N., Merget, D., & Schroeder, R. (2006). The effect of behavioral realism and form realism of real-time avatar faces on verbal disclosure, nonverbal disclosure, emotion recognition, and copresence in dyadic interaction. *Presence: Teleoperators and Virtual Environments*, *15*(4), 359-372.

[26] Kramer, N. C., Bente, G. & Piesk, J. (2003). The ghost in the machine: The influence of embodied conversational agents on user expectations and user behaviour in a TV/VCR application. In G. Bieber & T. Kirste (eds)., IMC Workshop 2003, Assistance, Mobility, Applications (pp. 121-128).

[27] Vardoulakis, L. (2013). Social Desirability Bias and Engagement in Systems Designed for Long-Term Health Tracking. Doctoral dissertation. Northwestern University.

[28] Fannon, K. (2003). 'Needle-stick'- A role-play simulation: Transformative learning in complex dynamic social systems. *International Journal of Training Research, 1,* 100-116.

7. NOTES

1. In accordance with Brooke's scoring scheme [22], participants' totals for the System Usability Scale were converted to a scale that ranges from 0 to 100.

2. All other effects failed to approach significance, Fs < 0.67, ps > .41. Eleven participants failed to complete the FNE, and thus were not included in this ANOVA analysis.

3. The interaction effect was also not significant for system usability (SU) scores, $F(1, 141) = 0.004$, $p = .95$, $d = 0.01$.

4. This work was supported by DARPA under contract W911NF-04-D-0005 and the U.S. Army. Any opinion, content or information presented does not necessarily reflect the position or the policy of the United States Government, and no official endorsement should be inferred.