2019-11

Innovative Accounting Interviewing: A Comparison of Real and Virtual Accounting Interviewers

Matthew D. Pickard
Northern Illinois University

Ryan Schuetzler
Brigham Young University

Joseph S. Valacich
University of Arizona

David A. Wood
Brigham Young University, davidwood@byu.edu

Follow this and additional works at: https://scholarsarchive.byu.edu/facpub

Part of the Business Commons

Innovative Accounting Interviewing: A Comparison of Real and Virtual Accounting Interviewers

Matthew D. Pickard
Northern Illinois University
mpickard@niu.edu

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

Joseph S. Valacich
University of Arizona
valacich@arizona.edu

David A. Wood
Brigham Young University
davidwood@byu.edu

November 2019

Innovative Accounting Interviewing: A Comparison of Real and Virtual Accounting Interviewers

Abstract

Recent technological advances have made it possible to create automated, virtual interviewers, called embodied conversational agents (ECAs). We study how an ECA compares to a human interviewer in three experiments. In experiment 1, we show that two theoretically motivated factors—making the ECA facially and vocally similar to the interviewee—result in the ECA performing similarly to or better than human interviewers for six antecedents of disclosure quality. In two additional experiments, we show that employees are on average between 21 and 32 percent more likely to disclose violating internal controls to an ECA than to a human interviewer, even if the human interviewer has significant interviewing experience. These findings contribute to the ECA design literature by showing that similarity-enhancing features of ECAs increase the antecedents of disclosure. The findings also contribute to the accounting literature by demonstrating that ECA technology can increase the scope of interviewing in accounting without reducing interview quality.

I. INTRODUCTION

The interview (i.e., verbal inquiry) is prevalent in accounting practice, especially in external and internal audit settings (Messier et al., 2017; Trompeter and Wright, 2010). The purpose of an audit interview is to obtain from individuals both "financial and non-financial" information that is new or that "differs significantly from other information" already gathered (International Auditing and Assurance Standards Board, 2009, para. A23). For example, auditors would benefit from capturing data about the "tone at the top" using a wide net to tease out insights about attitudes and practices in management. Similarly, although interviewing cannot be the sole source of evidence when testing controls, larger-scale interviewing would benefit both internal and external auditors seeking to understand the extent of employee compliance with internal controls and policies.

Although accountants interview in a variety of domains, evidence suggests that accountants struggle to interview effectively (Emett and Wood, 2010; Hall, 2005; Myers, 2006; McGimsey and Whelan, 2015) and that even experts can struggle to perform interviews well (Bond and DePaulo, 2006). Poor interviewing performance results in suboptimal information for decision-making. Interviewing is also time-consuming and costly. One possible way to reduce the cost and to increase the scalability of interviews is to automate part of the interviewing process. Automating interviews is viable for auditing only if the quality of the evidence gathered is high. In this study, we conducted initial tests of how interviewees responded to automated computer interviewers—referred to as embodied conversational agents (ECAs)—relative to human interviewers to understand how ECAs influence evidence quality. In their basic form, ECAs are computers with an animated visual representation. In their advanced form, ECAs are

autonomous and can employ emotion and other nonverbal communication to perform their tasks (Beale and Creed, 2009; Bickmore and Cassell, 2001). Figure 1 shows ECA image examples.

We expect ECAs to perform well in interviews that are relatively structured, have closed-ended questions, and do not require significant follow-up or probing questions. ECAs are also valuable where large sample sizes, which are costly to collect using human interviewers, are desirable. Information-gathering interviews often fit these characteristics. For example, SAS No. 99 requires external auditors to interview management, the audit committee, internal auditors, and others within the organization about the existence and suspicion of fraud. SAS No. 99 interviews include questions that are structured and close-ended, such as "Are you aware of any allegations of fraud?" and "Do you have any knowledge of fraud or suspected fraud affecting the entity?" While the ECA is unlikely to be effective at follow-up questions in these situations, the original data collection effort could be enhanced by using an ECA to identify where follow-up would be necessary. ECAs would currently not be well suited for admission-seeking interviews in fraud investigations. Such interviews are highly dynamic and require higher intelligence to navigate toward the end goal.

Prior ECA literature has found that ECAs perform as well as or better than humans in eliciting self-disclosure or in closely related dependent variables, such as impression management (Bailenson et al., 2006; DeVault et al., 2014; Gratch et al., 2014; Hasler et al., 2013; Kang and Gratch, 2010; Lind et al., 2013; Lucas et al., 2014; Pickard et al., 2016; von der Pütten et al., 2010; von der Pütten et al., 2011; Sproull et al., 1996). Although this literature suggests ECAs can be effective in areas outside accounting, empirical evidence within the accounting domain is still necessary for two reasons.

First, there is doubt in the accounting community about computer-mediated communications. For instance, there are many who believe face-to-face communication is irreplaceable in auditing contexts. This sentiment was captured by Kachelmeier (2018), who stated that "In contrast to any form of [written or technology-enhanced] communication, only the very old-fashioned medium of direct conversation can capture the nonverbal cues, gestures, and inflexions that add context and meaning to our words . . . face-to-face interaction [is] especially important in an audit environment for which success or failure depends on the effectiveness of interactions between audit and client personnel" (p. 58). Second, we found organizations were unwilling to test automated interviewing in their environment without sufficient evidence that ECAs would be effective in the accounting domain (see discussion and interview results in the online appendix). This was true even after we described findings from prior ECA studies that were conducted in other domains. Thus, applying research to an accounting setting will help influence practice to a greater degree by allowing organizations to see more direct evidence of research results in their working domain.

This study also contributes to the ECA literature by examining how ECA-interviewee facial and vocal similarity can increase information disclosure. We expected, based on the similarity-attraction theory from social psychology, that greater ECA-interviewee similarity leads to greater information disclosure. To test this hypothesis, we conducted an experiment where we induced ECA-interviewee similarity by morphing participants' faces into the ECA and using human voices rather than computer-generated voices.

To the authors' knowledge, only Vugt, Bailenson, Hoorn, and Konijn (2010) had previously manipulated ECA facial similarity, but their dependent variables were involvement and usefulness, not information disclosure. Likewise, studies that have manipulated the ECA's

voice measured its influence on variables other than information disclosure, such as social presence, consumer trust, and persuasion (Lee, 2010; Pickard et al., 2012; Qiu and Benbasat, 2005a, 2005b). Therefore, we contribute to the ECA literature by studying how ECA facial and vocal similarity influence information disclosure.

We conducted three experiments to understand the effect of ECA-interviewee similarity and to compare ECA interviewers to human interviewers. The between-subject, experimental results of study 1 show that either a human voice or a similar face increases perceived similarity and that perceived similarity is associated with an increase in six antecedents of disclosure quality (social attraction, social presence, power, trustworthiness, expertise, and likability). The results also show that the ECA conditions designed to enhance similarity performed as well as or better than the human condition for most of the antecedents of information disclosure. Although the results show that increasing the similarity of the ECA through facial similarity and voice manipulations increases antecedents of disclosure quality, we did not find statistically significant associations between the antecedents of disclosure quality and some measures of disclosure quality—likely because of poor measures of disclosure quality. Thus, we designed a second and third experiment in a real-world accounting setting to better measure information disclosure.

In the second and third between-subject experiments, we had human and ECA interviewers ask university employees about their compliance with timekeeping internal controls and policies. To provide greater internal and external validity, we varied the experience level of the human interviewers and allowed the human interviewers more latitude in using disclosure-enhancing tactics in these experiments. In the third experiment, we added an additional condition to see how human and ECA interviewers compare to an internal control questionnaire.

In both the second and third experiments, the employees reported more internal control violations to the ECA than they did to the human interviewers or in the internal control questionnaire. The results show meaningful effect sizes. In experiment 2, employees reported violating internal controls 65.6 percent of the time to the ECA versus 43.9 percent of the time to human interviewers, a statistically and economically significant difference of 21.1 percent. In experiment 3, the employees reported violating internal controls 54.1 percent of the time to the ECA and only 21.7 percent of the time to the human interviewers and 32.6 percent of the time in the internal control questionnaire—both statistically significant differences from the ECA. Notably, the use of more experienced interviewers in the third experiment did not lead to increased disclosure.

This paper makes important contributions to the ECA literature, accounting research, and accounting practice. We inform ECA literature by demonstrating the effectiveness of the similarity-attraction principle in ECA design to increase the antecedents of information disclosure. Our model can help future researchers and practitioners design ECAs that are likely to elicit greater information disclosure.

This paper also finds that ECAs can elicit more information disclosure in accounting-related interviews than human interviewers and internal control questionnaires. The current study suggests that ECAs perform better than humans in structured interviews that have relatively closed-ended questions and do not require significant follow-up or probing questions. Thus, ECAs may be helpful for the following type of audits: American Institute of Certified Public Accountants (AICPA) System and Organization Controls (SOC) type 1 and type 2 audits, International Organization for Standardization (ISO) audits (e.g., ISO 9000 series audits), European Union General Data Protection Regulation (GDPR) audits, Health Insurance

Portability and Accountability Act (HIPAA) audits, Payment Card Industry Data Security Standard (PCI-DSS) audits, Financial Industry Regulatory Authority (FINRA) audits, and Federal Information Security Management Act (FISMA) audits. In addition, ECAs could be useful in operational audits, whistle-blower hotline audits, p-card audits, and tax audits. All these types of audits use, at least in part, structured, closed-ended questions.

Given the cost, reach, and scalability advantages of ECAs over humans, these results should interest accounting organizations that conduct interviews. This is the first paper to test the ability of automated interviewers to elicit information disclosure in accounting interviews. Therefore, it demonstrates the need for additional research to explore how to optimize the design of ECAs to automate accounting-related interviews.

II. LITERATURE REVIEW AND HYPOTHESES

Interviews are a valuable part of accounting practice. For instance, interviews can help expose misstatements in financial reporting (Hylas and Ashton, 1982; Wright and Ashton, 1989) and improve the reliability of evidence gathered (Liu 2012; Saiewitz and Kida, 2018; Walsh and Bull, 2010). Prior research has also sought to improve interview quality, studying such things as enhancing interview preparation (Liu, 2012) and simulating interview experiences (Buckless et al., 2014). Despite the importance of interviewing in accounting, some research suggests that interviewing in accounting could still be improved (Canadian Institute of Chartered Accountants 2000; Emett and Wood, 2010; Hall, 2005; McGimsey and Whelan, 2015; Myers, 2006). One possible way to improve interviewing in accounting is to use ECAs.

ECA research is rooted in the "computers are social actors" (CASA) paradigm. CASA posits that people often treat computers as social beings (Nass and Moon, 2000). For example, people demonstrate politeness and gender differences in their communications with computers

(Nass et al., 1994) and people feel increased liking, trust, and submissiveness towards a computer that shows them empathy (Beale and Creed, 2009; Brave et al., 2005).

CASA has driven the application of social psychology theories to human-computer interactions. For example, supporting similarity-attraction theory, research demonstrates that users prefer to interact with computers similar to themselves (Nass et al., 1995; Nass et al., 1996; Nass and Moon, 2000). The literature repeatedly demonstrates that computers do not have to engage in complex conversational behaviors for communication theories and norms to hold in human-computer interactions. Thus, simple ECAs with static, scripted dialogues may be usefully employed to automate many accounting interviews and to provide several potential advantages over human interviewers.

The first potential advantage of an ECA over a human interviewer is that ECAs are naturally objective and unbiased, meaning that the ECA will ask questions exactly as it is programmed without being influenced by mood, emotion, fatigue, and other human factors. Accountants can lack objectivity and are biased in some settings (e.g., see Bazerman et al., 2002; Hogarth, 1991; Kahneman et al., 1982; Smith and Kida, 1991). Since human interviewers often unintentionally communicate biases nonverbally to the interviewee, ECAs may improve the quality of information gathered in interviews simply because the interviewee does not perceive immediate ulterior motives.

A second advantage of ECA interviewers is their potential to elicit greater information disclosure in some situations (Lind et al., 2013; Lucas et al., 2014; Pickard et al., 2016). Evidence suggests that virtual agent characteristics influence how honest people are with their virtual interviewers (Schuetzler et al., 2018). Interviewees may feel less apprehensive disclosing

sensitive information to an ECA because they perceive ECAs as not capable of social or emotional judgments (Pickard et al., 2016).

Other advantages of ECA interviewers relate to cost-savings. ECA interviewers do not become fatigued, are scalable, and can simultaneously interview dozens of interviewees across the world, increasing the reach of interviews while avoiding some human resource and travel costs. The ability to simultaneously interview multiple people can have important implications for fraud investigations (Brody et al., 2015).

Compared to humans, ECAs also pose some disadvantages. To date, ECAs have limited abilities to ask appropriate follow-up questions, to interpret nonverbal cues, and to detect emotions. This limits their ability to fully engage in and adapt to complex, real-time conversations. Thus, human interviewers are currently irreplaceable when the interview requires significant probing and follow-up (e.g., admission-seeking fraud interviews). This study marks the beginning of a research stream to determine when and where ECAs can effectively augment accounting interviews.

To begin this research stream, we turn to prior research on similarity-attraction theory. This theory suggests that perceived similarity increases the persuasive power of individuals (Burger et al., 2004; Jiang et al., 2010; Silvia, 2005). Studies have shown that similarity increases the antecedents of disclosure (Bailenson, Isyengar et al., 2008; Chartrand and Bargh, 1999; Lakin et al., 2003; Posthuma et al., 2002; Montoya et al., 2008; Murphy and Strong, 1972). Therefore, we expect individuals to disclose more information to interviewers whom they perceive to be more similar to themselves. While there are many facets of similarity (e.g., attitudes, personalities, backgrounds), we focus on the effects of physical appearance and voice, as both are highly salient and easily manipulated in an ECA.

One's physical appearance is arguably best defined by one's face. One's face is also a feature one is most familiar with and the main feature in which one identifies oneself. Therefore, facial similarity is a powerful characteristic to activate the similarity-liking effect. Bailenson, Isyengar et al. (2008) increased voters' preferences for presidential candidates by creating photos that blended the voters' faces into the candidates' faces—making the candidates look more similar to the voters. Mackinnon, Jordan, and Wilson (2011) showed that people prefer to sit with physically similar strangers rather than with physically different strangers. Outside human-to-human relationships, Roy and Christenfeld (2004) showed that humans are likely to pick pets that resemble them.

The voice of an ECA is another salient and easily manipulated ECA characteristic. Individuals rate voices that sound similar to their own as more likable and are more willing to engage in telephone surveys with people who have voices similar to their own (Fern et al., 1986; Oksenberg et al., 1986; der Vaart et al., 2006). There is some evidence that computerized voices can reduce individuals' social responses to computers (Lee, 2008, 2010). Therefore, we expect interviewees to perceive natural human voices as more similar to their own than computer-generated voices.

In summary, we expected a similar-looking and similar-sounding ECA to encourage interviewees to disclose higher-quality information compared to a different-looking and different-sounding ECA. This theory does not explicitly describe the exact relation between perceived similarity and increased disclosure for all levels of perceived similarity—that is, the theory does not clarify whether ever-increasing levels of perceived similarity will continue to increase disclosure. Thus, we had no grounds for predicting whether our two manipulations of similarity would interact to increase disclosure more than just the additive effect of each

manipulation. Therefore, we did not make an interaction prediction for the two similarity-enhancing manipulations, but we did test for a possible interaction effect in our models. We formalized only the following hypotheses:

H1: Interviewees will perceive an interviewer to be more similar to themselves if the interviewer has a face designed to be similar to that of the interviewee.

H2: Interviewees will perceive an interviewer to be more similar to themselves if the interviewer uses a human voice rather than a computer-generated voice.

H3: Increases in perceived interviewer-interviewee similarity will lead to higher-quality information disclosure.

Finally, the purpose of designing and testing an ECA to increase disclosure was to compare it to a human interviewer. This general comparison allows the research and practice community to respond to the prospect of using ECA interviewers in the field. If the ECA were worse than the human, future studies could improve the design of the ECA and accounting practice could wait to see whether promising results are obtained in future studies. If the ECA were better than the human, accounting practice could begin pilot studies using the ECA in real-world environments while researchers could study all the reasons why ECAs are superior.[1] Taking this latter approach could increase accounting relevance to practice and facilitate a more complete understanding of a complex phenomenon.[2] We therefore explored how our ECAs compare to human interviewers in the following research question:

---

[1] Providing a general test is also useful for informing standard setters, so they can consider changing standards, or at least for providing guidance to encourage auditors to try alternative methodologies. As discussed by Christ et al. (2019), standard setters and audit personnel often have significantly different views of what is permitted by standards related to technology, and increased discussion and encouragement by standard setters are likely to increase innovation in auditing.

[2] As noted by Kachelmeier (2018) in discussing Bennett and Hatfield's (2018) study of how written versus email communications influence auditors' professional skepticism, "Experimentalists generally take one step at a time. There is good reason for this approach, as the ability to isolate causal influences under ceteris paribus conditions is the key comparative benefit of a controlled experiment. Still, for something as complex as auditor-client interaction, the conclusions reached by any single-step experiment are virtually guaranteed to be incomplete" (p. 61). As discussed in the next section, we tried to take both approaches. We conducted an experiment in which we studied

III. RESEARCH METHODOLOGY

We conducted three experiments. In experiment 1, we tested our three hypotheses and explored our research question. In experiments 2 and 3, we continued to explore the research question and extend experiment 1 by (1) testing in a real-world accounting setting, (2) testing measures of an accounting phenomenon (internal control violations), (3) testing a precise measure of information disclosure (admissions of internal control violations), (4) relaxing several of the artificial constraints used in experiment 1, and (5) testing an alternative information-gathering methodology (using internal control questionnaires instead of interviews).

Experiment 1

To test our hypotheses and explore our research question, we conducted a laboratory experiment with a 2 (dissimilar vs. similar ECA face) x 2 (computer-generated vs. human voice) + 1 (human interviewer control group) between-subjects factorial design.

Participants

In total, 290 participants completed the first experiment. We recruited participants from undergraduate business classes at a large, public university in the southwestern United States. We offered participants extra credit for their participation. As shown in the demographic statistics in Table 1, participants represented a broad range of ethnicities, relatively dispersed ages, and both genders.[3]

---

one theoretical factor, perceived similarity, for understanding the relation between ECAs and interviewee disclosure, and then we conducted additional experiments that relaxed artificial experimental constraints to allow for a more robust testing of the research question in a more real-world setting.

[3] There were not any significant differences in demographic variables between conditions and including demographic variables in our analyses did not influence the results.

We designed the task to be something with which students would be familiar: performing a job interview (in this case it was with the FBI).[4] To accommodate space constraints in the lab and to handle the experimental manipulations, we conducted the human interviewer control group several weeks before the ECA treatment groups.

Human Interview Procedures

At least a week before arriving for the study, participants completed a pre-experiment survey that collected demographic information. Upon arriving at the laboratory and after consenting to participation, participants were reminded that they would be interviewed for an intelligence analyst position with the FBI. Because internal security is important to the FBI, they would be asked some personal questions. The participants were given the interview questions (see Appendix 1) and were instructed to take a few minutes to formulate answers to each question. This gave the participants time to recall the necessary information and formulate their responses and helped ensure that non-disclosure was by choice and not due to lack of recall.

When the participants were ready, we directed them to a private area. We informed them that the interviewer would not repeat any questions, ask follow-up questions, or answer any questions they might ask. This ensured that we treated all participants equally and helped control the differences between the human and ECA treatments. We then emphasized that the interviewer would wait patiently for them to respond after asking each question.

The human interviews were conducted through Skype video-conference calls.[5] The gender-matching interviewer sat in a separate room in front of a web camera. The interviewer wore a black t-shirt and a black baseball cap with "FBI" embroidered on the front (see Figure 1).

---

[4] Though the job interview with the FBI was hypothetical, we intentionally included questions in the interview that would elicit sensitive, personal information if answered honestly (see Appendix A for the questions we asked).

[5] As one of the advantages of ECAs—and computer technology in general—is scalability, we chose to compare the ECA to a Skype video-conference call because Skype provides scalability for the human resource.

The experiment proctor initiated the video-conference call after the participant was seated. The interviewer answered the Skype call and waited for the participant to hold up his or her participant ID written on a piece of paper. The interviewer then introduced himself or herself, explained the interview process, and proceeded with the interview script and questions.

The participants held up a "Next Question" sign after they had completed their response to each question. They were told this would help to ensure that the interviewer did not interrupt their response before they completed it. In reality, the interviewer's speaker volume was muted, so the "Next Question" sign served as a signal for the interviewer to know when to proceed.[6] After the interview was finished, the interviewer thanked the participant and reminded him or her to complete the post-experiment survey.

These design choices come with costs, as they resulted in the human interviewers being more scripted than in a typical human interview. However, for the initial test, we attempted to make the human and ECA interview situations as similar to one another as possible to maintain internal validity. That is, we wanted to see how ECA and human interviewers were perceived while holding as many other things constant as possible. We relaxed several of these design choices in the second and third experiments.

ECA Interview Procedures

The participants interviewed by an ECA completed the pre-experiment survey and attended a pre-experiment session to have their photograph taken approximately a week before the experiment. We told participants that we needed the photographs to test the interviewing

---

[6] Muting the volume on the interviewer's side provided several benefits. First, it maximized the video-conference call's sound quality by eliminating audio feedback and echoes. Second, it minimized the possibility that the interviewers would react to the verbal responses of the participants. The human interviewers practiced maintaining a neutral demeanor before the experiment. These controls made the human interviewer situation more comparable to the ECA, which followed a preprogrammed, static interview script.

system's facial recognition capabilities. However, we used the photographs to create ECA models of those who were randomly assigned to the facial similarity treatments.[7]

The ECA interview followed the same script that the human interviewers used. After the participants were seated in front of a computer in a private area, the computer instructed them to enter their participant ID. The gender-matching ECA interviewer then appeared on the screen. After the proctor left the room, the participants started the interview by clicking a button at the bottom of the screen. When the participants completed their response to a question, they clicked a button at the bottom of the screen to prompt the ECA to ask the next question. After the participants responded to the last question, the computer automatically presented the post-experiment survey.

Experimental Manipulations and Variable Measurements

We manipulated human and computer-generated voices by recording the two human interviewers reading the interview script and by using software for generating computer voices, respectively.[8] To create the similar-looking ECAs, we used Singular Inversions' FaceGen Modeller to generate ECA models from the photos of the participants taken in the pre-experiment session. Figure 1 shows an example of how the human interviewers appear as ECAs using this technology. For the ECA interviews in which facial similarity was tested, we used an ECA model generated from the participant's photo as the interviewer. For the other ECA interviews, we used the ECAs generated for the male or female human interviewer presented in Figure 1.

---

[7] We view this slight element of deception as necessary in order to maintain experimental control and to be able to manipulate the ECAs without the participant's knowledge. We note that the Institute Review Board who reviewed this protocol had no concerns with this small amount of deception.

[8] We used Loquendo's™ TTS audio generation software to create the computer-generated voices. In both the human- and computer-generated-voice conditions, we used Visage Technologies'™ Real-Time Automatic Lip Sync module to sync the ECA's lips to the audio files.

To test our hypotheses and research question, we needed a way to measure information disclosure. Prior research illustrates that measuring information disclosure is challenging. Most measures of self-disclosure are trait-based scales that measure an individual's general tendency to be open with personal information (Collins and Miller, 1994; Cozby, 1973). However, a state-based measurement is needed to compare interviewer effectiveness. State-based disclosure has been measured in a variety of ways (e.g., total number of words spoken, the sensitivity of the message contents, and the frequency of personal references); however, each of these has its limitations.[9] Because of disparities between state-based disclosure measurements, Goodstein and Reinecker (1974) argue that it is not possible to draw conclusions about which is superior.

To deal with this challenge, in the first study, we focused most of our attention on studying the antecedents of information disclosure quality. Given prior research showing that these antecedents lead to more disclosure, we expected that if we could find a relation between perceived similarity and antecedents of disclosure, this would lead to more disclosure in naturally occurring settings. In experiments 2 and 3, we focused more directly on disclosure by measuring admissions to internal control and policy violations.

We relied on persuasion research to select six antecedents of information disclosure (see Petty and Wegner (1998) for a detailed review of these variables in the persuasion literature).[10] In essence, when interviewers successfully elicit disclosure from interviewees, the interviewers are persuading the interviewees to share information. The specific antecedents of information disclosure that we examined are social attraction, social presence, power, trustworthiness,

---

[9] For example, a response may contain many words but reveal very little. According to Omarzu (2000), the risk perceived in disclosing information has a direct and inverse effect on the sensitivity of a disclosure. To measure the sensitivity of a response, either the interviewees must self-report the amount of sensitive information they revealed or external raters must evaluate the sensitivity post-hoc. Since sensitivity is subjective, both approaches are problematic.

[10] The specific antecedents were chosen based on discussion with a leading researcher in the communications field and based on a search of the literature.

expertise, and likability.[11] Prior research has found links between each of these constructs and disclosure (Skalski and Tamborini, 2007).

Social attraction is "a motivational state in which an individual is predisposed to think, feel, and usually behave in a positive manner toward another person" (Simpson and Harris, 1994, p. 47). Prior research has shown a reciprocal relationship between attraction and disclosure: individuals disclose more to those to whom they are attracted, and they are attracted more to those to whom they disclose more (Collins and Miller 1994; Taylor and Hinds 1985). Moon (2000) states, "The disclosure-attraction relationship [is] as robust as any found in the psychological literature" (p. 324).

Social presence "refers to the perceived degree to which a communication medium conveys interactants' presence. Presence involves perceptions of both intimacy and immediacy and operates as a function of the availability of verbal, nonverbal, and contextual cues in that medium" (Ruppel et al., 2017, p. 20). In simpler terms social presence is "the sense of being with another" (Biocca et al., 2003, pg. 457). As reported in a meta-analysis by Ruppel et al. (2017), higher social presence is correlated with greater disclosure.

In relationships, power is defined as "the ability or capacity to change another person's thoughts, feelings, or behavior so they align with one's own desired preferences." (Simpson et al., 2015, p. 409). Research shows that powerful people elicit more information than they provide to less powerful people (Goffman, 1956; Slobin et al., 1968).

Rotter (1967) defines trustworthiness as "an expectancy held by an individual or group that the word, promise, verbal or written statement of another individual or group can be relied

---

[11] The social presence measure is not reviewed in Petty and Wegner (1998), but it is included in other literature in this domain (see, e.g., Fogg, 2002; Skalski and Tamborini, 2007).

upon" (p. 651). As expected, research has shown that people disclose more to those whom they trust (Metzger, 2007; Steel, 1991).

Sternberg and Sternberg (2012) define expertise as "superior skills or achievement reflecting a well-developed and well-organized knowledge base" (p. 468). Accounting research shows that more expertise on the audit committee and/or board of directors is associated with more voluntary disclosure in companies (Akhtaruddin and Haron, 2010; Gul and Leung, 2004).

Finally, likability is a positive affective evaluation of an object (Roskos-Ewoldsen and Fazio, 1992). A meta-analysis conducted by Collins and Miller (1994) of the relation between self-disclosure and liking showed that "liking others leads us to disclose more to them" (p. 468).

We measured the antecedents of disclosure quality following well-established scales in prior research. Full citations for each scale are included in Table 2. In testing the antecedents of disclosure quality, we do not provide an in-depth discussion or testing of all the possible nuances that exist between these different antecedents, perceived similarity, and disclosure quality. For example, it is possible that some of these antecedents are more or less important than others, that an ECA may have a greater effect on some, etc. We were unable to test all possible questions related to these issues in this single study; rather, the purpose of this paper is to provide initial evidence about how ECAs and human interviewers compare, and we encourage future researchers to continue to analyze and to improve our understanding of these topics.[12]

We attempted to measure overall disclosure quality by treating *Disclosure Quality* as a formative construct. To measure this construct, we captured several different variables to

---

[12] Research needs to examine each individual antecedent and to evaluate all the antecedents together to provide a comprehensive view of this phenomenon. As the reader will readily recognize, this presents a chicken-and-egg scenario of deciding with which approach to start. We chose a general comparison to demonstrate the importance of ECAs and to encourage future research to continue studying the many additional open questions in this line of research.

measure the amount of information disclosed, the truthfulness of information disclosed, and the ease with which it was disclosed. Following existing literature, we measured *TotalWords* (Collins and Miller, 1994; Joinson, 2001; Strassberg and Anchor, 1975;), *FirstPerson* (Collins and Miller, 1994; Yang et al., 2017), *LexicalDiversity* (Burgoon and Hale, 1983; DePaulo et al., 2003; Ekman, 2009; Zhou et al., 2004), and *DisfluencyRatio* (Kang and Gratch, 2010)*.* In addition, we measured *InterjectionRatio*, *DifficultyAnswering*, and *HonestyAnswering* to see whether they help to measure this construct. We calculated the linguistic measures (i.e., *TotalWords*, *LexicalDiversity*, *DisfluencyRatio*, and *InterjectionRatio*) for only the responses to the open-ended and sensitive questions (specifically, questions 3, 4, 6, 10, 11, 12, and 14). We list the specifics of how all the variables are measured in Table 2. Our hypotheses suggest that increased perceived similarity will increase disclosure, so we expected that perceived similarity should also increase the antecedents of disclosure (with the exception of social presence), which in turn influence disclosure quality. We present this model in Figure 2.

As seen in Panel B of Table 3, following prior research, we retained measures of the antecedents of disclosure quality only if the outer loadings for these measures were greater than 0.70 (Hair, Jr. et al., 2016; Henseler et al., 2012). Also, we show that the Cronbach Alpha's, composite reliability, and average variance extracted for these measures are all above acceptable levels of 0.70, 0.70, 0.50, respectively (see discussion in Garson [2012]). This supports the construct validity of our reflective indicator measures.

We did find some problems with discriminant validity in the model, meaning there is a high correlation between some of the constructs (Hampton, 2015). Specifically, we found that the HTMT criterion is above 0.90 between *Expertise* and both *Power* and *Trustworthy* (Henseler et al., 2014). To confirm that this is not problematic, we (1) dropped each of the troubled

constructs from the model and in a separate analysis (2) included only each construct without any other constructs and confirmed that in both cases, the results are qualitatively similar to those reported. Thus, although there is high correlation between some of the constructs, it does not appear to impact any of the statistical inferences that are made.

In terms of the formative measure of *Disclosure Quality*, we retained four of the seven possible measures based on face validity and analysis of the path loading significance (Garson, 2012). That is, we retained the measures that conceptually should increase or decrease disclosure quality that had the expected sign (e.g., *FirstPerson* and *HonestyAnswering* have positive relations with *Disclosure Quality*, and *LexicalDiversity* and *DifficultyAnswering* have negative relations with *Disclosure Quality*).[13] We note that none of the variables had statistically significant loadings. Overall, we believe our measure of *Disclosure Quality* is a relatively poor measure but that the antecedents of disclosure quality are strong measures. The relatively poor showing of *Disclosure Quality* is one reason for conducting experiments 2 and 3 with better disclosure measures.

Experiment 2

Participants

We collaborated with the internal audit group at a large, private university. The internal audit group identified entry-level employees from diverse jobs for testing. In total 129 part-time student employees participated. Demographic information about participants is listed in Table

---

[13] We note that our results are similar to the inclusion of any/all of the other three measures and the exclusion of any of the measures we retained. Results are also similar if we use only question responses that we deemed to be the most sensitive in nature (i.e., responses to questions 6, 11, 12, and 14).

4.[14] All participants were required to follow internal controls relative to how many hours they worked, clocking in for work, and not working on personal activities during work.

Interview Procedures

Upon arrival at the study, we randomly assigned participants to interview with either an ECA or one of three human interviewers. The interviews were conducted in a small room with the door closed. Interviewees were told that their responses would all be anonymized and that they would have no effect on their employment status unless we were required to disclose something by law.[15] The interview questions were developed in conjunction with the head of internal audit for the institution and can be seen in Appendix 2. Once the interview was completed, the interviewee was directed to a computer where he or she completed a questionnaire about the interview.

For this study, we used a male ECA with a human voice (see Figure 3) for all participants in the ECA condition. We hired three male Master of Accountancy students to be the human interviewers. We hired male interviewers to match the gender of the ECA. The head of internal audit determined that all three interviewers were hirable as part-time internal auditors at the institution based on their resumes. Furthermore, the head of internal audit assured us that these types of interview questions would be asked by someone with the experience level of our interviewers.

We instructed the human interviewers to maintain a neutral demeanor, to limit (but not entirely remove) any personal chitchat and similar activities, and to ask no follow-up questions.

---

[14] Demographic variables, other than gender, do not differ significantly between conditions. Gender differs between conditions at the p-value < 0.10 level. The results are robust to including demographic variables as control variables in the analyses.

[15] This requirement was mandated by the IRB of the participating institution.

These design choices were meant to keep the human interviewer relatively similar to the ECA but to make the human interviewer situation more realistic than in experiment 1.

Manipulations

To have high statistical power for our tests, we chose to test only a single ECA condition in the second experiment. Given the results from the first experiment (discussed in the next section), we chose to use the voice manipulation because it performed slightly better in some dimensions than the facial manipulation and is easier to implement in practice.

Model Variable Measurement

Our focus in the second experiment was to evaluate a direct measure of information disclosure: self-reporting of internal control violations. With the head of internal audit, we developed seven questions that asked about compliance with specific internal controls related to clocking in and working on approved tasks during business hours. These measures, which are defined in Table 2, include *WorkedOverLimit*, *SpreadTime*, *AskedWorkUncompensated*, *AskedSpreadTime*, and *ClockedInNotWorking*. We also included two general questions about being asked to do unethical things or to violate university policies—*AskedUnethical* and *AskedViolatePolicy*. We combine these seven questions into two overall measures, *ReportedICViolation* and *NumberOfICViolations*, which measure whether an interviewee disclosed any violation of internal controls and the number of internal control violations that were disclosed, respectively.[16]

---

[16] One of the authors coded the interview responses, as these were objective statements. To verify the coding quality, a colleague independent of the research paper and blind to the experimental hypotheses coded a random selection of 13 interviews. Based on the 10 relevant questions (i.e., questions 3 through 12 in Appendix 2), the independent coder agreed with the author on 99.2 percent of the codings (i.e., 129 out of the 130 codings).

Experiment 3

Experiment 3 used the same task as experiment 2, with slight modifications. We only highlight the differences between experiments 2 and 3 in this section.

Differences from Experiment 2

Working with the same internal audit group, we tested a new sample of 126 part-time student employees. All employees were required to follow the same internal controls tested in the previous study. Demographic information about participants is listed in Table 4.[17]

One limitation from study 2 is that the interviewers were inexperienced. In experiment 3, we employed five different interviewers who had significant accounting interviewing experience. All interviewers were deemed by the head of internal audit to be qualified to perform this task. The head of internal audit selected two current internal auditors employed by the university to conduct interviews: one with one year of experience and the other with three years of experience. We selected the other three interviewers from personal contacts: a controller who had seven years of controller or assistant controller experience, a former Big 4 audit manager with eight years of Big 4 audit experience and an additional 10 years of work experience, and a former Big 4 audit senior manager with nine years of Big 4 audit experience and an additional 23 years of work experience. All interviewers were male to match the gender of the ECA.

Another limitation in experiment 2 is that we may have limited the effectiveness of the human interviewers by instructing them to maintain a neutral demeanor, to limit any personal chitchat and similar activities, and to ask no follow-up questions. In experiment 3, we told the interviewers to do their best to collect complete and accurate information from the interviewee.

---

[17] Demographic variables do not differ significantly between conditions, and results are robust to including demographic variables as control variables in the analyses.

The only requirements we gave the interviewers were to read a required statement from the IRB and to get an answer to each of the questions.[18,19]

Finally, given the limited nature of the information requested in the internal control interview, it is possible that an internal control questionnaire could be used in place of an interview to gather the information from a broad group of people. Research has shown that self-administered surveys (i.e., questionnaires) can be a valuable and cost-effective way of gathering potentially sensitive information from many people (Lind et al., 2013). Thus, to further substantiate the effectiveness of ECA interviews, we added an additional comparison group that completed an internal control questionnaire instead of participating in an interview.

IV. RESULTS

Results for Experiment 1

We present the results of testing the research model (Figure 2) in Figure 4. Panel A of Table 3 provides coefficients, standard deviations, t-stats, and p-values for the model presented in Figure 4. To test H1–H3, we examined only the ECA conditions. To estimate the model, we used SmartPLS because the model contains both reflective and formative indicators (Sarstedt et al. 2016; Rigdon 2014).[20] To test H1 and H2, we examined the coefficients between *Similar Face* and *Similarity* and between *Human Voice* and *Similarity*, respectively. The results show

---

[18] Since the audio of all the interviews were recorded, we were able to observe interviewer behavior after the interviews were finished. From the voice recordings, we observed that the interviewers used various tactics including changing the order of the questions, changing the wording of questions, building rapport in various ways, asking follow-up questions, etc. Thus, it appears that the interviewers did their best to elicit complete and accurate information. We note that the tactics were not consistently applied, so it is not possible to examine the differential effectiveness of interview tactics.

[19] One of the authors coded the interview responses as these were objective statements. To verify the coding quality, a colleague independent to the research paper and blind to the experimental hypotheses coded a random selection of 14 interviews. Based on the 10 relevant questions (i.e., questions 3 through 12 in Appendix 2), the independent coder agreed with the author on 97.9 percent of the codings (i.e., 137 out of the 140 codings).

[20] We estimated bootstrapped samples using 500 subsamples. We found consistent results for all other bootstrapped sample sizes we tried.

that ECAs that either have a similar face or use a human voice produce greater perceptions of similarity (p-values < 0.05). Including both a similar face and a similar voice did not produce greater perceptions of similarity (p-value for interaction effect > 0.10). Thus, the evidence supports H1 and H2: using an ECA with a face similar to the interviewee's face or with a human voice increased perceptions of similarity.

To test H3, we examined all paths between *Similarity* and the six antecedents of disclosure quality and the relation between the six antecedents of disclosure quality and our proxy for disclosure quality. In each case, the effect between similarity and the antecedents of disclosure quality was statistically significant and in the direction expected (all p-values < 0.05).[21] We found no relation between the antecedents of disclosure quality and our measure of disclosure quality. This likely indicates that we had a poor measure of disclosure quality, which as discussed, is difficult to measure. Thus, our results partially support H3. That is, we found that perceived similarity increased antecedents of information disclosure but did not increase our poor proxy of information disclosure.[22]

To test our research question, we did two things. First, we used ANOVAs to compare the human condition to each of the ECA conditions using contrast coding. Second, we separately

---

[21] Panel D of Table 3 provides indirect effects for the *Similar Face*, *Human Voice*, and *Similarity* constructs. The results show that *Similar Face* and *Human Voice* had significant indirect effects on each of the antecedents of disclosure quality (p-values < 0.10). Although the disclosure quality measures were relatively poor, the *Similar Face* and *Human Voice* manipulations had significant effects on disclosure quality at the p-value = 0.166 and 0.156 levels, respectively.

[22] As an alternative way of testing the relation between our manipulations and the antecedents of disclosure quality, we removed the *Similarity* variable from the model (and excluded the interaction since it was not significant) and tested whether *Human Voice* and *Similar Face* directly affect the antecedents of disclosure quality. In testing this model, we found that *Human Voice* was positively and statistically associated with each antecedent of disclosure quality. We found that *Similar Face* was statistically related only to social presence at the p-value < 0.10 level. This suggests that using a human voice may be more effective at producing disclosure than using a face designed to be similar to the interviewee's face.

estimated the reduced model shown in Figure 5 for each experimental condition and then compared the coefficients in each model between the human and ECA conditions.

We provide descriptive statistics for all measures in Panel A of Table 5 and a correlation matrix in Table 6. The descriptive statistics show that in terms of perceived similarity, using either a similar face or a human voice produced comparable perceptions of similarity, as the only condition that differed from the other conditions was the non-similar face, computer-voice condition. The ANOVA comparisons in Table 5 show that the human condition performed significantly worse than all the ECA conditions in terms of *Social Attraction* but significantly better than all the ECA conditions in terms of *Power*. The human condition was better than the non-similar face, computer-generated voice condition for all the other antecedents of disclosure quality except *Likable* and was better than both the computer-generated voice conditions for *Expertise*. Notably, the similar face, human-voice ECA condition was significantly more likable than the human condition.

The significantly lower social attraction of the human could be a result of the unnatural scripted constraint we placed on the human interviewers. The interviewees expected them to behave more socially. Thus, the violated expectations regarding the interviewers' behavior may have rendered the interviewers less socially attractive. Similar reasoning could apply to the likability of the interviewers. However, the significantly greater likability of the similar-looking, human-voice ECA compared to the ECAs with a computer-generated voice provides evidence that an ECA can be designed to increase the antecedents of information disclosure.

The results showing the differences in path coefficients and associated p-values are shown in Table 7. This test allowed us to examine whether any path differs between the human condition and each of the ECA conditions—that is, whether the effect of perceived similarity had

the same effect on the antecedents of disclosure quality between the ECA and human conditions. The results show almost no statistically significant differences between groups (there are two marginally statistically significant differences between *Similarity* and *Power*). The data suggests that perceived similarity had virtually no statistically significant differential effect on the antecedents of disclosure quality between the human and ECA conditions.

Results for Experiment 2

We present experiment 2 results testing our research question in Table 8. Panel A shows that participants disclosed violating at least one internal control to an ECA 65.1 percent of the time compared to only 43.9 percent of the time when interviewed by a human—a statistically and economically significant difference (p-value < 0.05). Furthermore, participants were likely to disclose more violations of internal controls to an ECA than to a human (p-value < 0.05). In Panel B of Table 8, we separately analyze each type of internal control violation admission. We found that our results are driven by one internal control violation: working over the allowable number of hours.[23]

For this experiment, three different interviewers conducted the interviews, conducting 15, 23, and 28 interviews each. When we examined individual interviewer performance, we note that participants were more likely to disclose violating internal controls to the ECA (65.1 percent of the time) than to any of the interviewers (disclosure percentages were 20.0, 46.4, and 56.5 percent). The differences were statistically significant for only two of the interviewers (p-value <

---

[23] We note that the lack of differences for several of the internal control conditions may be the relative lack of participants who admitted to violating a more serious control to any interviewer—that is, a floor effect. The overall measures and the *WorkedOverLimit* measure were not subject to possible floor effects given their magnitude. Indeed, *WorkedOverLimit*, *ReportedICViolation,* and *NumberOfICViolations* were all statistically greater than zero (p-values < 0.001).

0.05). Thus, it appears that some interviewers can elicit more violations of internal controls than others, which represents an important area for future study.[24]

Results for Experiment 3

We present experiment 3 results testing our research question in Table 9. Panel A shows that participants disclosed violating at least one internal control to an ECA 54.1 percent of the time compared to only 21.7 percent of the time when interviewed by a human and 32.6 percent of the time when filling out an internal control questionnaire. Furthermore, participants disclosed more internal control violations to the ECA (0.70) than to the human interviewer (0.30) or on the internal control questionnaire (0.37). Simple comparisons show that the differences between the ECA and the human interviewers and between the ECA and questionnaire condition for both measures were statistically significant (all p-values < 0.05). Finally, there was no statistical difference between the human interviewer and questionnaire conditions. The magnitudes of the differences between the ECA and the human interviewers and between the ECA and questionnaire condition were meaningful, with more than twice as many violations disclosed to the ECA as to the human interviewer and sixty-five percent more disclosures to the ECA than in the questionnaire.

Panel B shows a multivariate comparison between conditions using a logistic regression. For this analysis, the ECA condition was included in the intercept so the coefficients on Human and Questionnaire show the difference between each condition and the ECA condition. The results are similar to the results of the univariate comparisons: the ECA was statistically better than the human condition at the p-value < 0.01 level and better than the questionnaire condition

---

[24] We did not conduct a similar breakdown of interviewers for experiment 1 to avoid confounding gender differences with interviewer differences since participants were matched based on their gender to either the female or male interviewer.

at the p-value < 0.05 level. Similarly, Panel C shows the ANOVA comparison for differences in *NumberOfICViolations* and shows the same pattern of results.

We also test to see if there are differences in interviewers in Experiment 3, since the interviewers had significantly different amounts and types of experience. We recognize that there are only a limited number of interviewers and that each interviewer conducted a relatively small number of interviews (ranging from 8 to 10 each), but we believe it is worth noting that the percent of internal control violation admissions for the five interviewers ranged only between 0.20 to 0.25 and these were not statistically different than each other (p-values > 0.80).[25] Indeed, there is no correlation with years of audit experience or years of business experience and being able to get employees to disclose more internal control violations (all p-values > 0.60).[26]

V. CONCLUSION

The purpose of this study was to explore ways to make ECAs more effective interviewers and to compare ECAs to human interviewers in accounting settings. In our first experiment, we showed that two similarity-enhancing changes to an ECA—using a human voice instead of a computer-generated voice and increasing the facial similarity of the ECA to the interviewee—positively influenced the antecedents of disclosure. There was no significant additional benefit of using both features together. In our second and third experiments, we used a direct measure of disclosure—admission of internal control violations—and found that interviewees were more willing to disclose violating internal controls to an ECA interviewer than to a human interviewer or on an internal control questionnaire. Taken together, our results provide robust evidence

---

[25] It is also worth noting that between experiment 2 and 3, the more experienced interviewers did not improve the interviewing of the less experienced interviewers who were required to follow the script word-for-word.

[26] After conducting both experiment 2 and 3 with the same internal audit group, we shared the results with the head of the internal audit group. He was surprised at how well the ECA did, especially against his internal auditors and the more experienced interviewers. The results were sufficiently convincing that we are in talks of how we can further develop the ECA to be deployed via the web, rather than requiring interviewees to physically go to a computer lab, for field testing with their organization.

across multiple experiments of potential interviewing quality gains from using ECAs to automate interviews.

This study makes several important contributions to research and practice. First, this study provides initial support for the feasibility of using ECAs to automate basic, data-gathering interviews. The use of ECAs has the potential to reduce the costs and to increase the scalability of accounting interviews. Given the general skepticism about whether technology can effectively replace face-to-face interaction (see discussion in the introduction), this paper provides an initial, but important, data point in favor of the effective use of technology. Second, the results demonstrate that, despite advances in text-to-speech technology and the advantages it provides by increasing flexibility and requiring less effort (e.g.., there is no need to record a human), using a real human voice for an ECA is still more effective at establishing similarity.

Third, our study demonstrates that an ECA interviewer morphed to look like the interviewee can increase the antecedents of disclosure quality. This is a capability that is obviously unique to embodied agents and cannot be replicated with human interviewers. In this study, measures were taken ahead of time to manipulate specific facial features of the conversational agent to match those of the interviewees. As technology advances, however, such a manipulation could be done in real-time. It should also be noted here that the similar-looking ECA with a human voice in Study 1 was perceived as significantly more likable than the human interviewers in Study 1. Given the amount of research that establishes the link between liking and self-disclosure (see Collins and Miller [1994]), this is an important contribution to the ECA literature.

Lastly, this study highlights the difficulty of assessing disclosure in response to open-ended and more subjective questions (see, e.g., experiment 1), especially when ground truth is

not available. The closed-ended, precise, and more objective questions in experiments 2 and 3 made it easier to detect greater disclosure to the ECA. This provides support for our assessment that ECA interviewing will be most effective in structured, closed-ended interviews with relatively straightforward response options. Future research should continue to explore novel approaches to measure and improve disclosure to open-ended questions.

This study and its results are subject to several limitations and provide a springboard for future research. This study acts only as initial evidence that ECAs have potential for use in accounting settings. There are many situational factors that might affect the amount of information disclosed to an ECA or a human interviewer: company culture, sensitivity of the topics, personality traits of the human interviewer (e.g., unassuming, overbearing, aggressive), and severity of the possible consequences associated with the interview topic are a few examples. More research is needed to understand the situations where ECAs are superior to human interviewers, and vice versa. We especially encourage field research to test this across diverse accounting settings and topics. Field research will further lift the restraints of experimental settings that limit the findings of this study. Perhaps one of the most important restraints of experiments is the message conveyed to the participants that the situation is hypothetical and, therefore, without consequences.

This study raises the important question of what other characteristics of an ECA are important in creating the perception of similarity and liking in the user. For instance, perceived similarity and liking may also be increased using linguistic style (Taylor and Thomas, 2008), facial emotion (Hess and Fischer, 2013; Hess et al., 2001), and nonverbal gestures and body movements (Bailenson, Yee et al., 2008; Chartrand and Lakin, 2013; Lakin and Chartrand, 2003). Future research should consider taking a systematic and incremental approach in

introducing additional capabilities to an ECA-based automated interviewing system to determine which capabilities improve the disclosure of information and, thereby, improve the quality of the information gathered in the interview.

Our results show that ECAs have promise for enhancing and expanding accounting interview possibilities in firms. ECAs performing a broad set of interviews will free human resources to follow up on possible problems or to conduct deeper investigations.

REFERENCES

Akhtaruddin, M., and H. Haron. 2010. Board ownership, audit committees' effectiveness and corporate voluntary disclosures. *Asian Review of Accounting* 18 (1): 68–82.

Bailenson, J. N., S. Iyengar, N. Yee, and N. A. Collins. 2008. Facial similarity between voters and candidates causes influence. *Public Opinion Quarterly* 72 (5): 935–961.

Bailenson, J. N., N. Yee, D. Merget, and R. Schroeder. 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 & Virtual Environments* 15 (4): 359–372.

Bailenson, J. N., N. Yee, K. Patel, and A. C. Beall. 2008. Detecting digital chameleons. *Computers in Human Behavior* 24 (1): 66–87.

Bazerman, M. H., G. Loewenstein, and D. A. Moore. 2002. Why good accountants do bad audits. *Harvard Business Review* 80 (11): 96–103.

Beale, R., and C. Creed. 2009. Affective interaction: How emotional agents affect users. *International Journal of Human-Computer Studies* 67 (9): 755–776.

Bennett, G. B., and R. C. Hatfield. 2018. Staff auditors' proclivity for computer-mediated communication with clients and its effect on skeptical behavior. *Accounting, Organizations and Society* 68–69 (July): 42–57.

Bickmore, T., and J. Cassell. 2001. Relational agents: A model and implementation of building user trust. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 396–403. Seattle, WA, United States: ACM.

Biocca, F., C. Harms, and J. K. Burgoon. 2003. Toward a more robust theory and measure of social presence: Review and suggested criteria. *Presence: Teleoperators & Virtual Environments* 12 (5): 456–480.

Bond, C. F., and B. M. DePaulo. 2006. Accuracy of deception judgments. *Personality and Social Psychology Review* 10 (3): 214.

Brave, S., C. Nass, and K. Hutchinson. 2005. Computers that care: Investigating the effects of orientation of emotion exhibited by an embodied computer agent. *International Journal of Human-Computer Studies* 62 (2): 161–178.

Brody, R. G., M. D. Pickard, and J. Agins. 2015. Avatar interviewers: Computer-assisted interviews: Fact or folly? *Fraud Magazine* 30 (4): 22–27.

Buckless, F. A., K. Krawczyk, and D. S. Showalter. 2014. Using virtual worlds to simulate real-world audit procedures. *Issues in Accounting Education* 29 (3): 389–418.

Burger, J. M., N. Messian, S. Patel, A. del Prado, and C. Anderson. 2004. What a coincidence! The effects of incidental similarity on compliance. *Personality and Social Psychology Bulletin* 30 (1): 35–43.

Burgoon, J. K., and J. L. Hale. 1983. Dimensions of communication reticence and their impact on verbal encoding. *Communication Quarterly* 31 (4): 302–312.

Canadian Institute of Chartered Accountants. 2000. *Audit Enquiry: Seeking More Reliable Evidence from Audit Enquiry*. Toronto, Canada.

Chartrand, T. L., and J. A. Bargh. 1999. The chameleon effect: The perception–behavior link and social interaction. *Journal of Personality and Social Psychology* 76 (6): 893–910.

Chartrand, T. L., and J. L. Lakin. 2013. The antecedents and consequences of human behavioral mimicry. *Annual Review of Psychology* 64 (1): 285–308.

Christ, M. H., S. A. Emett, S. L. Summers, and D. A. Wood. 2019. *Prepare for Takeoff:*

Improving Audit Efficiency and Effectiveness with Drone-Enabled Inventory Audit Procedures. Working paper, University of Georgia, Arizona State University, Brigham Young University, and Brigham Young University.

Collins, N. L., and L. C. Miller. 1994. Self-disclosure and liking: A meta-analytic review. *Psychological Bulletin* 116 (3): 457–475.

Cozby, P. C. 1973. Self-disclosure: A literature review. *Psychological Bulletin* 79 (2): 73–91.

DePaulo, B. M., J. J. Lindsay, B. E. Malone, L. Muhlenbruck, K. Charlton, and H. Cooper. 2003. Cues to deception. *Psychological Bulletin* 129 (1): 74–118.

DeVault, D., R. Artstein, G. Benn, T. Dey, E. Fast, A. Gainer, … L. Morency. 2014. SimSensei Kiosk : A virtual human interviewer for healthcare decision support. *International Conference on Autonomous Agents and Multi-Agent Systems* (January): 1061–1068.

Ekman, P. 2009. *Telling Lies: Clues to Deceit in the Marketplace, Politics, and Marriage.* New York, NY. W.W. Norton.

Emett, S., and D. A. Wood. 2010. Common question-evasion tactics. *Journal of Accountancy* 210 (1): 36–39.

Fern, E. F., K. B. Monroe, and R. A. Avila. 1986. Effectiveness of multiple request strategies: A synthesis of research results. *Journal of Marketing Research* 23 (2): 144–152.

Fogg, B. J. 2002. *Persuasive Technology: Using Computers to Change What We Think and Do.* San Francisco, CA: Morgan Kaufmann Publishers.

Garson, D. 2012. *Partial Least Squares: Regression and Path Modeling.* Asheboro, NC: Statistical Publishing Associates.

Goffman, E. 1956. The nature of deference and demeanor. *American Anthropologist* 58 (3): 473–502.

Goodstein, L. D., and V. M. Reinecker. 1974. Factors affecting self-disclosure: A review of the literature. *Progress in Experimental Personality Research* 7: 49.

Gratch, J., G. M. Lucas, A. A. King, and L.-P. Morency. 2014. It's only a computer: The impact of human-agent interaction in clinical interviews. In *Proceedings of the 2014 International Conference on Autonomous Agents and Multi-Agent Systems*, 85–92.

Gul, F. A., and S. Leung. 2004. Board leadership, outside directors' expertise and voluntary corporate disclosures. *Journal of Accounting and Public Policy* 23 (5): 351–379.

Hair Jr., J. F., G. T. M. Hult, C. Ringle, and M. Sarstedt. 2016. *A Primer on Partial Least Squares Structural Equation Modeling (PLS-SEM)*. Los Angeles. Sage Publications.

Hall, J. J. 2005. Answer please: Fraud-based interviewing. *Journal of Accountancy* 200 (2): 61–65.

Hampton, C. 2015. Estimating and reporting structural equation models with behavioral accounting data. *Behavioral Research in Accounting* 27 (2): 1–34.

Hasler, B. S., Tuchman, P., & Friedman, D. (2013). Virtual research assistants: Replacing human interviewers by automated avatars in virtual worlds. *Computers in Human Behavior*, 29(4), 1608–1616.

Henseler, J., C. M. Ringle, and M. Sarstedt. 2012. Using partial least squares path modeling in advertising research: Basic concepts and recent issues. In *Handbook of Research on International Advertising*, 252–278. Northhampton, MA: Edgar Elgar Publishing Limited.

Henseler, J., C. M. Ringle, and M. Sarstedt. 2014. A new criterion for assessing discriminant validity in variance-based structural equation modeling. *Journal of the Academy of Marketing Science* 43 (1): 115–135.

Hess, U., S. Blairy, and S. B. Ursula Hessa. 2001. Facial mimicry and emotional contagion to

dynamic emotional facial expressions and their influence on decoding accuracy. *International Journal of Psychophysiology* 40 (2): 129–141.

Hess, U., and A. Fischer. 2013. Emotional mimicry as social regulation. *Personality and Social Psychology Review* 17 (2): 142–157.

Hogarth, R. M. 1991. A Perspective on cognitive research in accounting. *The Accounting Review* 66 (2): 277–290.

Hylas, R. E., and R. H. Ashton. 1982. Audit detection of financial statement errors. *The Accounting Review* 42 (4): 751.

International Auditing and Assurance Standards Board. 2009. Audit Evidence. In *ISA No. 500*.

Jiang, L., J. Hoegg, D. W. Dahl, and A. Chattopadhyay. 2010. The persuasive role of incidental similarity on attitudes and purchase intentions in a sales context. *Journal of Consumer Research* 36 (5): 778–791.

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.

Joseph, W. B. 1977. Effect of communicator physical attractiveness and expertness on opinion change and information processing. The Ohio State University.

Kachelmeier, S. J. 2018. Testing auditor-client interactions without letting auditors and clients fully interact : Comments on Bennett and Hatfield (2018). *Accounting, Organizations and Society* 69 (July): 58–62.

Kahneman, D., P. Slovic, and A. Tversky. 1982. *Judgment under Uncertainty: Heuristics and Biases*. New York, NY: Cambridge University Press.

Kang, S. H., and J. Gratch. 2010. Virtual humans elicit socially anxious interactants' verbal self-disclosure. *Computer Animation and Virtual Worlds* 21 (3–4): 473–482.

Lakin, J. L., and T. L. Chartrand. 2003. Using nonconscious behavioral mimicry to create affiliation and rapport. *Psychological Science* 14 (4): 334–339.

Lakin, J. L., V. E. Jefferis, C. M. Cheng, and T. L. Chartrand. 2003. The chameleon effect as social glue: Evidence for the evolutionary significance of nonconscious mimicry. *Journal of Nonverbal Behavior* 27 (3): 145–161.

Lee, E. J. 2008. Flattery may get computers somewhere, sometimes: The moderating role of output modality, computer gender, and user gender. *International Journal of Human-Computer Studies* 66 (11): 789–800.

———. 2010. The more humanlike, the better? How speech type and users' cognitive style affect social responses to computers. *Computers in Human Behavior* 26 (4): 665–672.

Lind, L. H., M. F. Schober, F. G. Conrad, and H. Reichert. 2013. Why do survey respondents disclose more when computers ask the questions? *Public Opinion Quarterly* 77 (4): 888–935.

Liu, G. 2012. Gathering evidence through enquiry: A process improvement focus. *Behavioral Research in Accounting* 24 (2): 153–175.

Lucas, G. M., J. Gratch, A. King, and L.-P. Morency. 2014. It's only a computer: Virtual humans increase willingness to disclose. *Computers in Human Behavior* 37: 94–100.

Mackinnon, S. P., C. H. Jordan, and A. E. Wilson. 2011. Birds of a feather sit together: Physical similarity predicts seating choice. *Personality & Social Psychology Bulletin* 37 (7): 879–92.

McGimsey, C. L., and D. Whelan. 2015. Forensic interviews: Plan to succeed. *Journal of Accountancy* 220 (2): 34–42.

Messier, W. F., S. M. Glover, and D. F. Prawitt. 2017. *Auditing & Assurance Services: A Systematic Approach*. 10th ed. New York, NY: McGraw-Hill Education.

Metzger, M. J. 2007. Communication privacy management in electronic commerce. *Journal of Computer-Mediated Communication* 12 (2): 335–361.

Montoya, R. M., R. S. Horton, and J. Kirchner. 2008. Is actual similarity necessary for attraction? A meta-analysis of actual and perceived similarity. *Journal of Social and Personal Relationships* 25 (6): 889–922.

Moon, Y. 2000. Intimate exchanges: Using computers to elicit self-disclosure from consumers. *Journal of Consumer Research* 26 (4): 323–339.

Murphy, K. C., and S. R. Strong. 1972. Some effects of similarity self-disclosure. *Journal of Counseling Psychology* 19 (2): 121–124.

Myers, R. 2006. Interviewing techniques: Tips from the pros. *Journal of Accountancy* 202 (2): 53–55.

Nass, C., B. J. Fogg, and Y. Moon. 1996. Can computers be teammates? *International Journal of Human Computer Studies* 45 (6): 669–678.

Nass, C., and Y. Moon. 2000. Machines and mindlessness: Social responses to computers. *Journal of Social Issues* 56 (1): 81–103.

Nass, C., Y. Moon, B. J. Fogg, B. Reeves, and C. Dryer. 1995. Can computer personalities be human personalities? In *Conference Companion on Human Factors in Computing Systems*, 228–229.

Nass, C., J. Steuer, and E. R. Tauber. 1994. Computers are social actors. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems: Celebrating Interdependence*, 72–78.

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

Ohanian, R. 1990. Construction and Validation of a Scale to Measure Celebrity Endorsers' Perceived Expertise, Trustworthiness, and Attractiveness. *Journal of Advertising* 19 (3): 39–52.

Oksenberg, L., L. Coleman, and C. F. Cannell. 1986. Interviewers' voices and refusal rates in telephone surveys. *The Public Opinion Quarterly* 50 (1): 97–111.

Omarzu, J. 2000. A disclosure decision model: Determining how and when individuals will self-disclose. *Personality and Social Psychology Review* 4 (2): 174–185.

Petty, R. E., and D. T. Wegener. 1998. Attitude change: Multiple roles for persuasion variables. *The Handbook of Social Psychology* 1 (323): 320.

Pickard, M. D., J. L. Jenkins, and J. F., Jr. Nunamaker. 2012. Embodied Agents and the Predictive Elaboration Model of Persuasion--The Ability to Tailor Embodied Agents to Users' Need for Cognition. *Hawaii International Conference on System Sciences* 0: 543–552.

Pickard, M. D., C. A. Roster, and Y. Chen. 2016. Revealing sensitive information in personal interviews: Is self-disclosure easier with humans or avatars and under what conditions? *Computers in Human Behavior* 65 (December): 23–30.

Posthuma, R. A., F. P. Morgeson, and M. A. Campion. 2002. Beyond employment interview validity: A comprehensive narrative review of recent research and trends over time. *Personnel Psychology* 55 (1): 1–81.

von der Pütten, A. M., L. Hoffmann, J. Klatt, and N. C. Krämer. 2011. Quid pro quo? Reciprocal self-disclosure and communicative accomodation towards a virtual interviewer. In *Intelligent Virtual Agents. IVA 2011. Lecture Notes in Computer Science*, edited by H. H.

Vilhjálmsson, S. Kopp, S. Marsella, and K. R. Thórisson, 6895:183–194. Springer Berlin Heidelberg.

von der Pütten, A. M., N. C. Krämer, J. Gratch, and S. H. Kang. 2010. ''It doesn't matter what you are!" Explaining social effects of agents and avatars. *Computers in Human Behavior* 26: 1641–1650.

Qiu, L., and I. Benbasat. 2005a. Online consumer trust and live help interfaces: The effects of text-to-speech voice and three-dimensional avatars. *International Journal of Human-Computer Interaction* 19 (1): 75–94.

———. 2005b. An investigation into the effects of text-to-speech voice and 3D avatars on the perception of presence and flow of live help in electronic commerce. *ACM Transactions on Computer-Human Interaction (TOCHI)* 12 (4): 329–355.

———. 2009. Evaluating anthropomorphic product recommendation agents: A social relationship perspective to designing information systems. *Journal of Management Information Systems* 25 (4): 145–182.

Reysen, S. 2005. Construction of a new scale: The Reysen likability scale. *Social Behavior & Personality: An International Journal* 33 (2): 201–208.

Rigdon, E. E. 2014. Rethinking partial least squares path modeling: Breaking chains and forging ahead. *Long Range Planning* 47 (3): 161–167.

Roskos-Ewoldsen, D. R., and R. H. Fazio. 1992. The accessibility of source likability as a determinant of persuasion. *Personality and Social Psychology Bulletin* 18 (1): 19–25.

Rotter, J. B. 1967. A new scale for the measurement of interpersonal trust. *Journal of Personality* 35 (4): 651–665.

Roy, M. M., and N. J. S. S. Christenfeld. 2004. Do dogs resemble their owners ? *Psychological Science* 15 (5): 361–363.

Ruppel, E. K., C. Gross, A. Stoll, B. S. Peck, M. Allen, and S. Y. Kim. 2017. Reflecting on connecting: Meta-analysis of differences between computer-mediated and face-to-face self-disclosure. *Journal of Computer-Mediated Communication* 22 (1): 18–34.

Saiewitz, A., and T. Kida. 2018. The effects of an auditor's communication mode and professional tone on client responses to audit inquiries. *Accounting, Organizations and Society* 65 (Feb): 33–43.

Sarstedt, M., C. M. Ringle, and J. F. Hair. 2016. Partial least squares structural equation modeling. In *Handbook of Market Research*, edited by C. Homburg, M. Klarmann, and A. Vomberg. Heidlebert: Springer International Publishing.

Schuetzler, R. M., J. S. Giboney, G. M. Grimes, and J. F. Nunamaker. 2018. The influence of conversational agent embodiment and conversational relevance on socially desirable responding. *Decision Support Systems* 114 (May): 94–102.

Silvia, P. J. 2005. Deflecting reactance: The role of similarity in increasing compliance and reducing resistance. *Basic and Applied Social Psychology* 27 (3): 277–284.

Simpson, J. A., A. K. Farrell, M. M. Oriña, and A. J. Rothman. 2015. Power and social influence in relationships. In *APA Handbook of Personality and Social Psychology, Volume 3: Interpersonal Relations.*, edited by P. R. Mikulincer, S. J. A, and J. F. Dovidio, 393–420. Washington, DC: American Psychological Association.

Simpson, J. A., and B. A. Harris. 1994. Interpersonal attraction. In *Perspectives on Close Relationships*, edited by A. L. Weber and J. H. Harvey, 45–66. Needham Heights, MA: Allyn and Bacon.

Skalski, P., and R. Tamborini. 2007. The role of social presence in interactive agent-based

persuasion. *Media Psychology* 10 (3): 385–413.

Slobin, D. I., S. H. Miller, and L. W. Porter. 1968. Forms of address and social relations in a business organization. *Journal of Personality and Social Psychology* 8 (3): 289–293.

Smith, J. F., and T. Kida. 1991. Heuristics and biases: Expertise and task realism in auditing. *Psychological Bulletin* 109 (3): 472–489.

Sproull, L., M. Subramani, S. Kiesler, J. H. Walker, and K. Waters. 1996. When the interface is a face. *Human-Computer Interaction* 11 (2): 97–124.

Steel, J. L. 1991. Interpersonal correlates of trust and self-disclosure. *Psychological Reports* 68 (3_suppl): 1319–1320.

Sternberg, R. J., and K. Sternberg. 2012. *Cognitive Psychology*. 6th ed. Belmont, CA: Wadsworth, Cengage Learning.

Strassberg, D. S., and K. N. Anchor. 1975. Rating intimacy of self-disclosure. *Psychological Reports* 37 (2): 562.

Taylor, D. A., and M. Hinds. 1985. Disclosure reciprocity and liking as a function of gender and personalism. *Sex Roles* 12 (11–12): 1137–1153.

Taylor, P., and S. Thomas. 2008. Linguistic style matching and negotiation outcome. *Negotiation and Conflict Management Research* 1 (3): 263–281.

Trompeter, G., and A. Wright. 2010. The world has changed—Have analytical procedure practices? *Contemporary Accounting Research* 27 (2): 669–700.

der Vaart, W., Y. Ongena, A. Hoogendoorn, and W. Dijkstra. 2006. Do interviewers' voice characteristics influence cooperation rates in telephone surveys? *International Journal of Public Opinion Research* 18 (4): 488–499.

Vrij, A. 1994. The impact of information and setting on detection of deception by police detectives. *Journal of Nonverbal Behavior* 18 (2): 117–136.

Vugt, H. C. V, J. N. Bailenson, J. F. Hoorn, and E. A. Konijn. 2010. Effects of facial similarity on user responses to embodied agents. *ACM Transactions on Computer-Human Interaction (TOCHI)* 17 (2): 1–27.

Walsh, D., and R. Bull. 2010. What really is effective in interviews with suspects? A study comparing interviewing skills against interviewing outcomes. *Legal and Criminological Psychology* 15 (2): 305–321.

Wright, A., and R. H. Ashton. 1989. Identifying audit adjustments with attention-directing procedures. *The Accounting Review* 64 (4): 710.

Yang, D., Z. Yao, and R. Kraut. 2017. Self-disclosure and channel difference in online health support group. In *Proceedings of the 11th International Conference on Web and Social Media, ICWSM 2017*, 704–707.

Zhou, L., J. K. Burgoon, J. F. J. Nunamaker, and D. Twitchell. 2004. Automating linguistics-based cues for detecting deception in text-based asyncronous computer-mediated communication. *Group Decision and Negotiation* 13 (1): 81–106.

Appendix 1—Experiment 1 Interview Script

Introduction: Hi, I am Agent Jones. I will be interviewing you today for a position with the FBI. I have fifteen questions I need to ask you. For each question, please wait until I have finished asking the question before responding. When you are done responding, please raise your "Next Question" sign and I'll ask you the next question. Some of the questions will be hard to answer. It's okay for you to take your time while formulating your responses. I'll wait patiently. Here is the first question:

1. What is your country of citizenship?

2. How long have you lived in the United States?

3. What characteristics of yourself are you most proud of?

4. What has been the biggest disappointment in your life?

5. What language do you speak most fluently?

6. Tell me about a time when you disregarded or stretched an organizational policy or guideline in order to get something done.

7. Describe to me a time when you made a serious mistake. How did you handle the situation?

8. What country is stamped most often on your passport?

9. Have you ever or do you now hold a passport that was issued by a foreign government?

10. Think about people that really aggravate you. Why do they annoy or bother you?

11. What have you done in your life that you feel most guilty about?

12. If you found a wallet containing $1,000 and no identification in it, what would you do with it and why?

13. Have you ever had any account or credit card suspended or canceled for failing to pay as you agreed?

14. Tell me about a time when you told a serious lie to avoid a punishment.

15. Are you currently authorized to work in the United States?

16. What is your country of citizenship?

Conclusion: That concludes this portion of the interview. Please remember to complete the survey on your computer's desktop. Thank you for your time.

Appendix 2—Experiment 2 and Experiment 3 Interview Script

Human Introduction:
Hi, my name is *(Name)*. I am conducting an internal audit to assess *(University name's)* compliance with certain internal controls. I will ask you a series of questions about your employment. Please take your time to answer each question. If you would like me to repeat the question, just ask. If you cannot formulate a response or if you choose not to respond, please say "no response."

All of your responses will be kept confidential unless we are required to disclose them by law. The audio recordings from this interview will be deleted after your responses are transcribed. After their deletion, there will be no way to connect your responses back to you.

Let's begin the interview.

ECA Introduction:
Hi, my name is Matt. I am conducting an internal audit to assess *(University name's)* compliance with certain internal controls. I will ask you a series of questions about your employment. Please take your time to answer each question. If you would like me to repeat the question, click on the button that will appear in the upper-left. If you cannot formulate a response or if you choose not to respond, please say "no response." After finishing your response to any question, click the button that will appear in the upper-right to continue.

All of your responses will be kept confidential unless we are required to disclose them by law. The audio recordings from this interview will be deleted after your responses are transcribed. After their deletion, there will be no way to connect your responses back to you.

Let's begin the interview.

Questions in both treatments of the interview.
1. Please state your participant ID.
2. Please describe your job duties.
3. Are you an undergraduate or graduate student?
4. Are you an international student?
5. Typically, during the fall and winter semesters, undergraduate students are limited to working 20 hours per week and graduate students are limited to working 28 hours per week. Have you ever worked more than the appropriate number of hours in a week without prior approval?
6. If you ever worked more than the appropriate limit, did you report all of the hours you worked in that week, or did you spread your time to another week to avoid going over the limit?
7. Has your supervisor ever asked you to work uncompensated hours to avoid going over the 20 or 28-hour work limit? If so, what did you do?

8. Has your supervisor ever asked you to spread hours over more than one week to avoid going over a limit? If so, what did you do?
9. Are you required to clock in to record your working hours? If so, how do you clock in? For example, do you use a time clock on a wall, a Cisco phone on a desk, the Y App, a computer, etc.?
10. Have you ever been clocked in while not residing in your designated workspace or spent time while clocked in doing primarily personal activities? For example, being clocked in before walking or driving to campus or being clocked in and working on homework before focusing on your job-related duties. If so, please describe the situation.
11. Have you ever been asked by your supervisor to do something that made you feel uncomfortable or that was unethical? If so, please describe what it was and how you responded.
12. Have you ever been asked to perform a task that you thought or knew was in conflict with university policy? If so, please describe the circumstances.
13. Do you have any ideas or suggestions for how your department could operate more effectively to accomplish its primary mission and objectives?
14. Do you have any ideas or suggestions for how your department could operate more efficiently?
15. Is there anything else we should know about the culture or operations of your department?

That concludes our interview. Thank you for your time.

Figure 1

Illustration of the ECA Created from Photos of the Human Interviewers

Figure 2
Model Comparing the Effectiveness of Various Types of ECAs
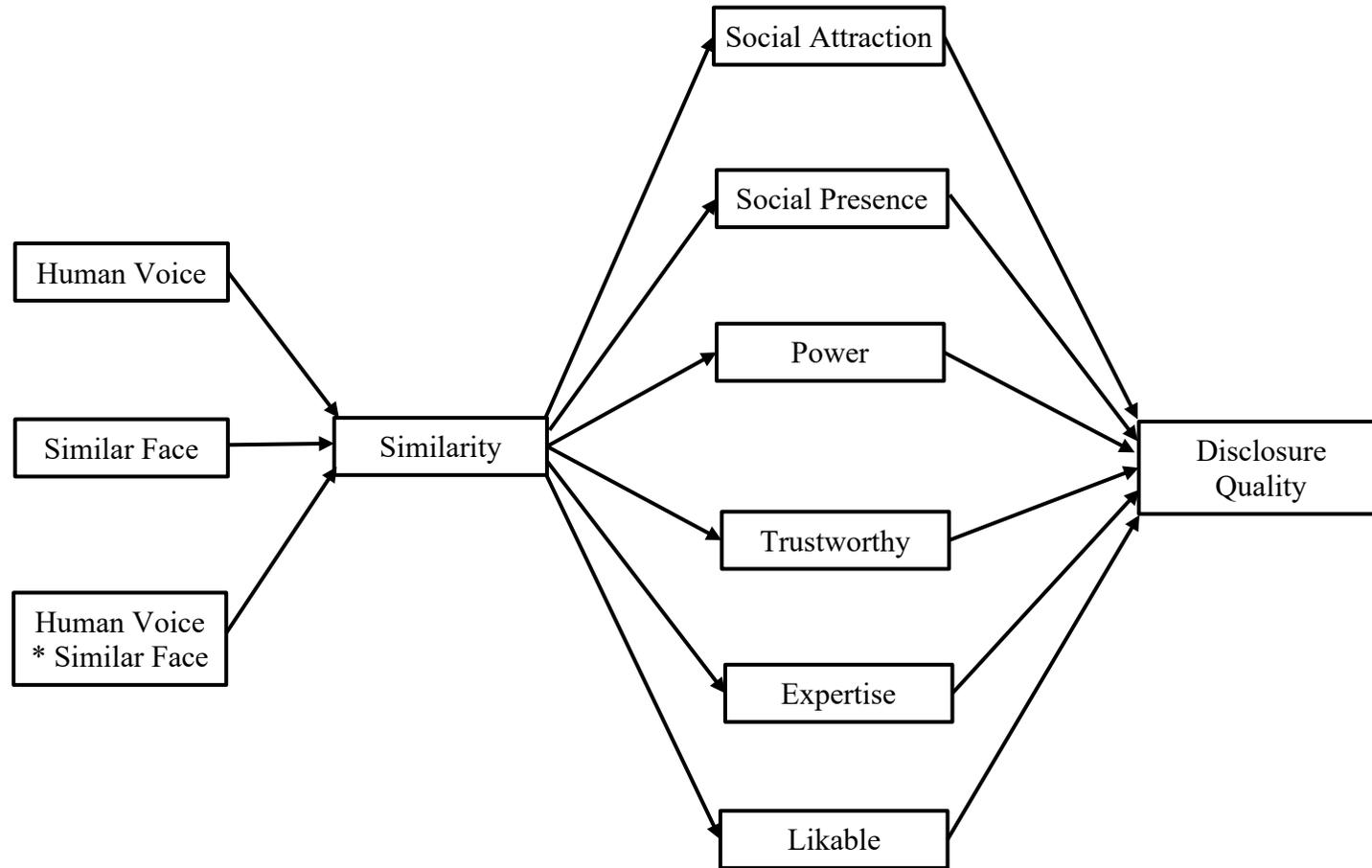
We expect that each of the paths will have a positive coefficient.

Figure 3
ECA Interviewer for Experiments 2 and 3

Figure 4
Model Comparing the Effectiveness of Various Types of ECAs
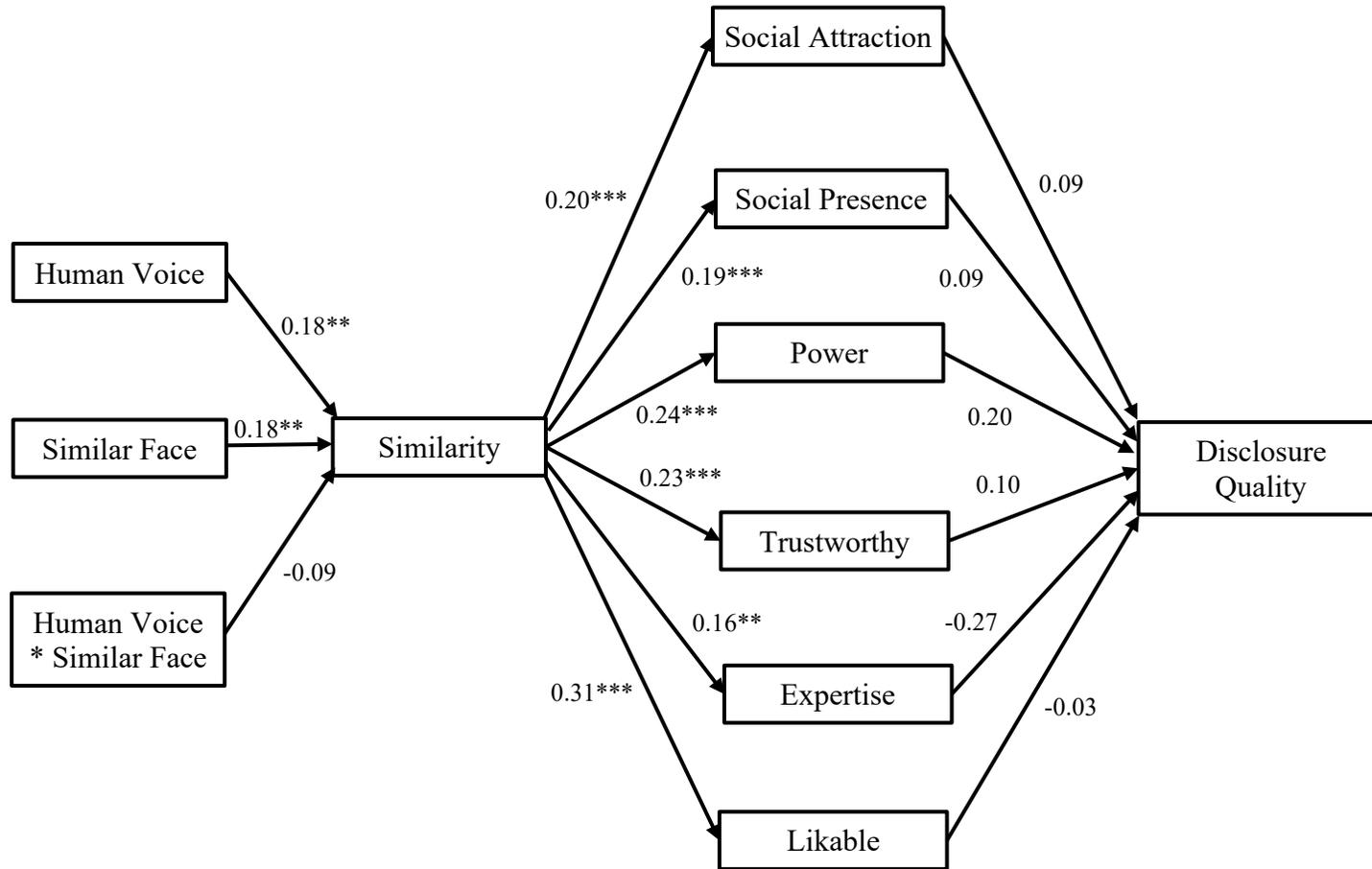
Numbers show path coefficients and ***, **, and * indicate p-values < 0.01, 0.05, and 0.10, respectively. Variables are defined in Table 1. Model fit indices are SRMR = 0.062 and NFI = 0.787.

Figure 5
Reduced Model for Comparing Human and ECA Interviewers
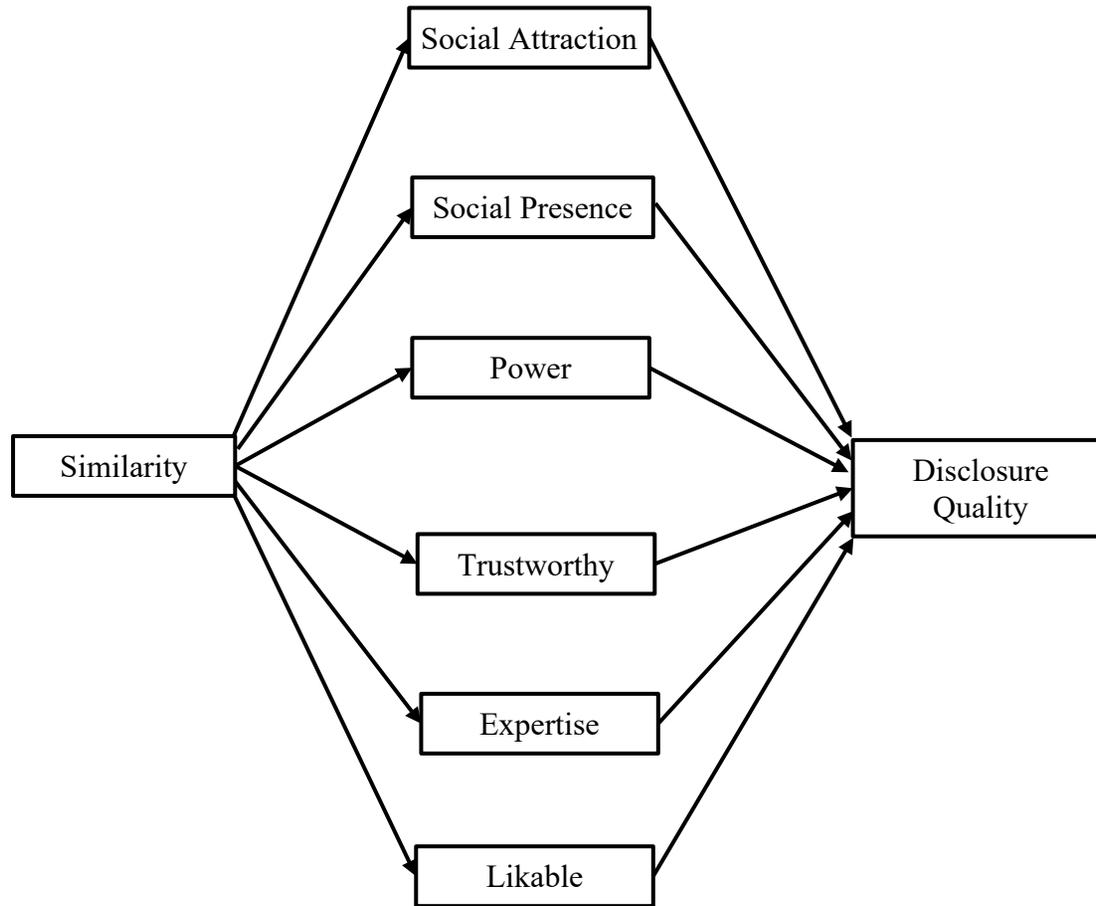
Table 1
Demographics of Participants in Experiment 1

Demographic	Human	Similar Face Human Voice	Similar Face Computer Voice	Non-Similar Face Human Voice	Non-Similar Face Computer Voice
Gender					
Female	29	27	28	24	30
Male	29	31	30	34	28
Age	20.97 (1.35)	22.21 (4.62)	22.09 (2.79)	21.26 (1.24)	22.83 (5.27)
Years of College Completed	2.74 (0.68)	2.80 (0.87)	2.74 (0.70)	2.85 (0.81)	2.94 (0.90)
GPA	3.49 (0.31)	3.43 (0.35)	3.39 (0.36)	3.37 (0.39)	3.41 (0.34)
Ethnicity					
White	45	39	43	38	42
Asian	6	10	8	12	6
Other	7	9	7	8	10
N	58	58	58	58	58

Counts or means and (stdev) are presented.

Table 2
Variable Definitions

Variable Name	Study	Definition
DifficultyAnswering	1	Participants self-reported the difficulty they had answering the 15 interview questions in experiment 1 (1 = very easy to 7 = very difficult). The 15 scores were averaged. The variable is reverse-coded so that higher values indicate greater disclosure quality (and easier time answering).
DisclosureQuality	1	Treated as a formative construct composed of TotalWords, FirstPerson, LexicalDiversity, DisfluencyRatio, InterjectionRatio, DifficultyAnswering, and HonestyAnswering.
DisfluencyRatio	1	The number of speech disfluencies divided by TotalWords. Disfluencies are breaks or irregularities that interrupt fluent speech. Examples include: huh, um, and like. Kang and Gratch (2010) used disfluency as a measure of disclosure.
Expertise	1	The participants were asked to what extent they agreed that the following adjectives described the interviewer (1 = strongly disagree to 7 = strongly agree): knowledgeable, qualified, skilled, informed, and competent. The scale was adapted from Nunamaker, Jr. et al. (2011) and Ohanian (1990).
FirstPerson	1	The number of first-person singular pronouns divided by TotalWords. Collins and Miller (1994) observed that disclosure is often "operationalized as . . . the number of self-relevant statements made during an interaction" (p. 458). Yang et al. (2017) used measure disclosure with first-person pronouns.
HonestyAnswering	1	Participants self-reported their honesty to the 15 interview questions in experiment 1 (1 = very dishonest to 7 very honest). The 15 scores were averaged.
InterjectionRatio	1	The number of speech interjections divided by TotalWords. An interjection is a part of speech that shows emotion. Examples include: eh, phew, and whoa.
LexicalDiversity	1	The number of unique words divided by TotalWords. Lexical diversity is theorized and used to measure deception in several studies (DePaulo et al., 2003; Ekman 2009; Vrij, 1994; Burgoon and Hale, 1983; Zhou

Variable Name	Study	Definition
		et al., 2004). Deception is the holding back or covering up of true information, so it is related to disclosure.
Likable	1	The participants were asked to what extent they agreed that the following adjectives described the interviewer (1 = strongly disagree to 7 = strongly agree): kind, cheerful, liking, friendly, good-natured, warm, nice, and attractive. The scale was adapted from Reysen (2005).
Power	1	The participants were asked to what extent they agreed that the following adjectives described the interviewer (1 = strongly disagree to 7 = strongly agree): power, dominant, confident, influential, high status. The scale was adapted from Nunamaker, Jr. et al. (2011) and Ohanian (1990).
Similarity	1	Self-reported assessment of how similar the interviewer was perceived to be to the interviewee. Measured on 7-point scale such that higher values indicate greater perceived similarity.
SocialAttraction	1	Four questions from Joseph (1977) designed to measure social attraction. The questions were "I think the interviewer could be a friend of mine," "It would be difficult to meet and talk with the interviewer," "The interviewer just wouldn't fit into my circle of friends," and "We could never establish a personal friendship with each other." Higher values indicate greater social attraction.
SocialPresence	1	Five items from Qiu and Benbasat (2009) designed to measure social presence. The items were "I felt a sense of human contact in the agent," "I felt a sense of personalness in the agent," "I felt a sense of human warmth in the agent," "I felt a sense of sociability in the agent," and "I felt a sense of human sensitivity in the agent." Higher values indicate greater social presence.
TotalWords	1	Total number of words spoken has been used as a measure of disclosure breadth in several studies (Strassberg and Anchor, 1975; Joinson, 2001). Collins and Miller (1994) also observe that disclosure is often "operationalized as the amount of time spent talking about oneself . . . during an interaction" (p. 458).

Variable Name	Study	Definition
Trustworthy	1	The trustworthiness factor of credibility. The participants were asked to what extent they agreed that the following adjectives described the interviewer (1 = strongly disagree to 7 = strongly agree): dependable, honest, reliable, sincere, and trustworthy. The scale was adapted from Nunamaker Jr. et al. (2011) and Ohanian (1990).
AskedSpreadTime	2,3	Participants were asked: "Has your supervisor ever asked you to spread hours over more than one week to avoid going over a limit?" It is coded 1 if they answered yes, and 0 if no.
AskedUnethical	2,3	Participants were asked: "Have you ever been asked by your supervisor to do something that made you feel uncomfortable or that was unethical?" It is coded 1 if they answered yes, and 0 if no.
AskedViolatePolicy	2,3	Participants were asked: "Have you ever been asked to perform a task that you thought or knew was in conflict with university policy?" It is coded 1 if they answered yes, and 0 if no.
AskedWorkUncompensated	2,3	Participants were asked: "Has your supervisor ever asked you to work uncompensated hours to avoid going over the 20- or 28-hour work limit?" It is coded 1 if they answered yes, and 0 if no.
ClockedInNotWorking	2,3	Participants were asked: "Have you ever been clocked in while not residing in your designated workspace or spent time while clocked in doing primarily personal activities?" It is coded 1 if they answered yes, and 0 if no.
NumberOfICViolations	2,3	The total number of internal control violations the participant reported throughout the entire interview.
ReportedICViolation	2,3	Coded 1 if the participant reported any internal control violation at any time during the interview, 0 otherwise.
SpreadTime	2,3	Participants were asked: "Has your supervisor ever asked you to spread hours over more than one week to avoid going over a limit?" It is coded 1 if they answered yes, and 0 if no.

Variable Name	Study	Definition
WorkedOverLimit	2,3	Participants were asked: "Have you ever worked more than the appropriate number of hours in a week without prior approval?" It is coded 1 if they answered yes, and 0 if no.

Table 3
Results of Testing the Model of Effectiveness of Various Types of ECAs

Panel A: Path Coefficients

Path Tested	Coefficient	Stdev	t-stat
Human Voice -> Similarity	0.18	0.09	2.10**
Similar Face -> Similarity	0.18	0.09	1.95**
Human Voice * Similar Face -> Similarity	-0.09	0.12	0.80
Similarity -> Social Attraction	0.20	0.07	2.79***
Similarity -> Social Presence	0.19	0.08	2.54***
Similarity -> Power	0.24	0.07	3.42***
Similarity -> Trustworthy	0.23	0.07	3.19***
Similarity -> Expertise	0.16	0.07	2.17**
Similarity -> Likable	0.31	0.07	4.54***
Social Attraction -> Disclosure Quality	0.09	0.11	0.83
Social Presence -> Disclosure Quality	0.09	0.18	0.49
Power -> Disclosure Quality	0.20	0.22	0.90
Trustworthy -> Disclosure Quality	0.10	0.24	0.39
Expertise -> Disclosure Quality	-0.27	0.29	0.95
Likable -> Disclosure Quality	-0.03	0.19	0.14

Panel B: Outer Loadings for Model's Reflective Indicators

	Expertise	Likable	Power	Social Attraction	Social Presence	Trustworthy
Bright	0.77					
Competent	0.85					
Expert	0.84					
Knowledgeable	0.89					
Qualified	0.90					
Skillful	0.87					
Trained	0.81					
Cheerful		0.85				
Friendly		0.85				
GoodNatured		0.82				
Kind		0.86				
Likable		0.85				
Nice		0.86				
Similar		0.74				
Warm		0.86				
Confident			0.79			
Dominant			0.73			
High Status			0.85			
Influential			0.82			
Powerful			0.83			
MeetAndTalk				0.79		

CircleOfFriends	0.80		
PersonalFriendship	0.83		
HumanContact		0.85	
Personalness		0.74	
HumanWarmth		0.90	
Sociability		0.90	
HumanSensitivity		0.85	
Dependable			0.83
Honesty			0.81
Reliable			0.84
Sincere			0.80
Trustworthy			0.84
Virtuous			0.75

Panel C: Construct Reliability and Validity Statistics for Model Reflective Indicators

Variable	Cronbach's Alpha	Composite Reliability	Average Variance Extracted
Social Attraction	0.74	0.85	0.65
Social Presence	0.90	0.93	0.72
Power	0.87	0.90	0.65
Trustworthy	0.90	0.92	0.66
Expertise	0.94	0.95	0.72
Likable	0.94	0.95	0.69

Panel D: Effect Size and Statistical Significance of Indirect Paths

Paths	Mean Effect	Stdev	T-stat	p-value
Human Voice -> Disclosure Quality	0.01	0.01	0.97	0.166
Human Voice -> Expertise	0.03	0.02	1.38	0.085
Human Voice -> Likable	0.06	0.03	1.87	0.031
Human Voice -> Powerful	0.05	0.03	1.66	0.049
Human Voice -> Social Attraction	0.04	0.02	1.58	0.058
Human Voice -> Social Presence	0.04	0.02	1.55	0.062
Human Voice -> Trustworthy	0.04	0.03	1.61	0.055
Similar Face -> Disclosure Quality	0.01	0.01	1.01	0.156
Similar Face -> Expertise	0.03	0.02	1.4	0.081
Similar Face -> Likable	0.06	0.03	1.84	0.033
Similar Face -> Powerful	0.04	0.03	1.69	0.046
Similar Face -> Social Attraction	0.04	0.02	1.53	0.063
Similar Face -> Social Presence	0.04	0.02	1.55	0.062
Similar Face -> Trustworthy	0.04	0.03	1.64	0.052
Similarity -> Disclosure Quality	0.05	0.05	1.13	0.131

***, **, and * indicate p-values < 0.01, 0.05, and 0.10, respectively. Variables are defined in Table 2.

Table 4
Demographic Statistics for Experiment 2 and Experiment 3

	Experiment 2		Experiment 3		
	ECA	**Human**	**ECA**	**Human**	**Questionnaire**
Gender (Female)	46.0% (50.2%)	30.3% (46.3%)	51.3% (50.7%)	50.0% (50.6%)	44.2% (50.2%)
Age	22.0 (1.6)	22.2 (3.1)	21.9 (1.6)	21.3 (2.0)	21.5 (2.0)
Hours worked per week	14.5 (5.2)	14.3 (5.2)	13.9 (4.3)	14.1 (5.4)	14.6 (5.6)
Program of Study					
Accounting	19	19	0	4	4
Business or Pre-Business	17	17	18	20	17
Finance	6	4	3	0	2
Information Systems	6	6	7	8	4
Marketing	6	6	0	3	7
Strategy	2	2	1	2	5
Other or Undecided	7	12	8	9	4
N	63	66	37	46	43

This table presents demographic information about all participants in experiment 2 and experiment 3. Not all participants answered every question, so results are presented for all responding participants for each question. For experiment 2, 92 unique job titles were listed. For experiment 3, 91 unique job titles were listed.

Table 5
Descriptive Statistics of Variables Measured in Experiment 1

Panel A: Descriptive Statistics, Mean (Stdev), N = 58 in all conditions

Variable	A Human	B Similar Face Human Voice	C Similar Face Computer Voice	D Non-Similar Face Human Voice	E Non-Similar Face Computer Voice
Similarity	$2.31\ (0.85)^{E}$	$2.39\ (1.03)^{E}$	$2.24\ (0.99)^{E}$	$2.24\ (0.93)^{E}$	$1.88\ (0.93)^{A,B,C,D}$
DisclosureQuality	1.76 (1.25)	1.71 (0.97)	1.83 (1.35)	1.62 (0.85)	1.58 (0.82)
SocialAttraction	$8.42\ (2.53)^{B,C,D,E}$	$9.93\ (2.87)^{A,E}$	$10.78\ (3.86)^{A}$	$10.12\ (3.18)^{A,E}$	$11.36\ (3.11)^{A,B,D}$
SocialPresence	$9.50\ (3.04)^{E}$	$10.29\ (3.67)^{C,E}$	$8.69\ (3.20)^{B}$	9.30 (3.25)	$8.16\ (2.77)^{A,B}$
Power	$18.88\ (3.87)^{B,C,D,E}$	$16.72\ (5.05)^{A,C}$	$14.36\ (4.45)^{A,B,D}$	$16.78\ (4.78)^{A,C}$	$15.27\ (5.45)^{A}$
Trustworthy	$20.61\ (4.10)^{E}$	$21.72\ (5.57)^{C,E}$	$18.83\ (6.47)^{B}$	$20.76\ (5.81)^{E}$	$18.50\ (6.17)^{A,B,D}$
Expertise	$28.93\ (4.85)^{C,E}$	$27.81\ (7.40)^{C}$	$25.10\ (7.62)^{A,B}$	26.84 (7.52)	$25.29\ (8.31)^{A}$
Likable	$23.04\ (5.40)^{B}$	$26.29\ (7.94)^{A,C,E}$	$22.31\ (8.16)^{B}$	$24.57\ (7.79)^{E}$	$21.03\ (7.67)^{B,D}$

Panel B: Comparison of Measures of Disclosure Quality, Mean (Stdev)

Variable	A Human	B Similar Face Human Voice	C Similar Face Computer Voice	D Non-Similar Face Human Voice	E Non-Similar Face Computer Voice
DifficultyAnswering	2.67 (0.63)	$2.85\ (0.54)^{C}$	$2.61\ (0.66)^{B}$	2.79 (0.54)	2.76 (0.52)
FirstPerson	0.10 (0.02)	$0.10\ (0.02)^{E}$	0.09 (0.02)	0.10 (0.02)	$0.09\ (0.02)^{B}$
HonestyAnswering	1.78 (1.35)	1.72 (1.04)	1.85 (1.45)	1.62 (0.91)	1.58 (0.88)
LexicalDiversity	0.48 (0.10)	0.51 (0.11)	0.51 (0.12)	0.52 (0.09)	0.51 (0.11)

Letters indicate statistically significant differences (p-values < 0.05) between conditions (as labeled with letters at the top of each column). Statistical comparisons are the results of testing contrasts in ANOVA for each different variable. Variables are defined in Table 2.

Table 6
Correlation Matrix for Experiment 1

	Similarity	DisclosureQuality	SocialAttraction	SocialPresence	Power	Trustworthy	Expertise
DisclosureQuality	0.06						
SocialAttraction	**-0.18**	-0.04					
SocialPresence	**0.18**	0.06	**-0.22**				
Power	**0.21**	0.03	**-0.17**	**0.40**			
Trustworthy	**0.19**	0.01	**-0.21**	**0.47**	**0.74**		
Expertise	**0.15**	-0.04	**-0.18**	**0.38**	**0.84**	**0.84**	
Likable	**0.27**	0.03	**-0.27**	**0.63**	**0.55**	**0.76**	**0.58**

Pearson correlation coefficients are presented. **Bold** (*italics*) signifies coefficient is significant at p-value < 0.05 (*0.10*). Variables are defined in Table 2.

Table 7
Comparison of Human and ECA Models

Variable	Differences in Path Coefficients between Human Condition and . . .			
	Similar Face Human Voice	Similar Face Computer Voice	Non-Similar Face Human Voice	Non-Similar Face Computer Voice
Similarity -> Social Attraction	0.28	0.22	0.12	0.21
Similarity -> Social Presence	0.11	0.21	0.18	0.04
Similarity -> Power	0.39*	0.37*	0.21	0.39
Similarity -> Trustworthy	0.24	0.13	0.06	0.12
Similarity -> Expertise	0.18	0.14	0.19	0.24
Similarity -> Likable	0.08	0.12	0.26	0.12
Social Attraction -> Disclosure Quality	0.04	0.41	0.44	0.58
Social Presence -> Disclosure Quality	0.16	0.48	0.23	0.48
Power -> Disclosure Quality	0.25	0.59	1.01	0.13
Trustworthy -> Disclosure Quality	0.70	0.82	0.04	0.40
Expertise -> Disclosure Quality	0.13	0.86	1.19	0.05
Likable -> Disclosure Quality	0.21	0.75	0.39	0.18

Reported numbers are differences in path coefficients (e.g., computed as Similar Face Human Voice condition minus human condition). ***, **, and * indicate p-values < 0.01, 0.05, and 0.10, respectively. Variables are defined in Table 2. These results test each of the different conditions against each other for the model in Figure 3.

Table 8
Analysis of Violation of Internal Controls Experiment 2

Panel A: Analysis of Reported Internal Control Violations, Mean (Stdev)

Variable	ECA	Human	Test Statistic
ReportedICViolation	65.1% (48.1%)	43.9% (50.0%)	$\chi^2 = 5.80$**
NumberOfICViolations	0.95 (0.92)	0.61 (0.89)	$t = 2.45$**
N	63	66	

Panel B: Analysis of Type of Reported Internal Control Violation, Mean (Stdev)

Variable	ECA	Human	χ^2
WorkedOverLimit	28.6% (45.5%)	13.6% (34.6%)	4.34**
SpreadTime	8.1% (27.5%)	4.5% (21.0%)	0.68
AskedWorkUncompensated	0.0% (0.0%)	1.5% (12.3%)	0.96
AskedSpreadTime	6.3% (24.6%)	4.5% (21.0%)	0.20
ClockedInNotWorking	17.7% (38.5%)	15.2% (36.1%)	0.16
AskedUnethical	0.0% (0.0%)	0.0% (0.0%)	0.00
AskedViolatePolicy	1.6% (12.6%)	0.0% (0.0%)	1.06

***, **, and * indicate p-values < 0.01, 0.05, and 0.10, respectively. Variables are defined in Table 2.

Table 9
Analysis of Violation of Internal Controls Experiment 3

Panel A: Analysis of Reported Internal Control Violations, Mean (Stdev)

Variable	ECA	Human	Questionnaire	Test Stat ECA v Human	Test Stat ECA v. Questionnaire	Test Stat Human v. Questionnaire
ReportedICViolation	54.1% (50.5%)	21.7% (41.7%)	32.6% (47.4%)	$\chi^2 = 9.28$***	$\chi^2 = 3.76$**	$\chi^2 = 1.32$
NumberOfICViolations	0.70 (0.78)	0.30 (0.66)	0.37 (0.58)	$t = 2.52$***	$t = 2.18$**	$t = -0.51$
N	37	46	43			

Panel B: Logistic Regression Comparing Conditions on Reported Internal Control Violations (ECA is included in intercept)

Variable	Estimate	Standard Error	χ^2
Intercept	0.163	0.330	0.24
Human	-1.444	0.486	8.81***
Questionnaire	-0.891	0.463	3.70**

Panel C: ANOVA Comparing Conditions on Number of Reported Internal Control Violations (ECA condition is comparison group)

Variable	Sum of Squares	F-value
Human	3.25	7.21***
Questionnaire	2.17	4.82**

***, **, and * indicate p-values < 0.01, 0.05, and 0.10, respectively. Variables are defined in Table 2.