Available online at www.sciencedirect.com
ScienceDirect
Computers in Human Behavior 23 (2007) 2175–2189
Computers in Human Behavior
www.elsevier.com/locate/comphumbeh

Social reactions toward people vs. computers: How mere lables shape interactions

E. Aharoni *, A.J. Fridlund

Department of Psychology, University of California, Santa Barbara, CA 93106, USA

Available online 18 April 2006

Abstract

What criteria afford a machine the status of a social agent? In this investigation, the mere label identifying an oral interviewer as human or computer was sufficient to affect participants' responses toward the interviewer during an online interview for a competitive mock job. Participants' impressions of the interviewer and self-reported emotional reactions to the interview were unaffected by the interviewer's identity. Despite this invariance, however, participants exhibited more interpersonal displays when the interviewer was identified as human. Overall, these results show that participants engaged in heightened impression management strategies (deferral to, or attempts to engage or appease) with the "human" interviewer. The computer interviewer did not merit equivalent social status.
© 2006 Elsevier Ltd. All rights reserved.

1. Introduction

Adults have been shown to apply the same social norms and rules of etiquette toward computers as they do toward other humans (De Laere, Lundgren, & Howe, 1998; Lautenschlager & Flaherty, 1990; Nass, Moon, Fogg, Reeves, & Dryer, 1995; Nass, Steuer, Henriksen, & Dryer, 1994; Reeves & Nass, 1996; Wenger, 1991). For instance, they will evaluate a computer tutor more favorably when in its presence than not, providing evidence for politeness (Nass, Moon, & Carney, 1999), and this can occur even without conscious awareness (Nass & Moon, 2000).

* Corresponding author. Tel.: +1 805 564 4638; fax: +1 805 893 4303.
E-mail address: aharoni@psych.ucsb.edu (E. Aharoni).

0747-5632/$ - see front matter © 2006 Elsevier Ltd. All rights reserved.
doi:10.1016/j.chb.2006.02.019

How does our interaction with machines compare with normal social commerce? Many of these studies of Social Reactions to Communication Technology (SRCT) did not supplement their observations with a human control group, making direct comparison impossible. Studies that have employed a human control group have tended to show that computer stimuli do not meet the standard for human interaction, eliciting considerably weaker interpersonal behavior than do human stimuli (Bonito, Burgoon, & Bengtsson, 1999; Lee & Nass, 2002; Martin & Nagao, 1989; See Parise, Kiesler, Sproull, & Waters, 1999, for an exception). This appears to be true even when the features of the stimuli are identical, and only the identity label is changed. For example, in immersive virtual environments, anthropomorphic virtual agents elicit a higher threshold of social influence when they are believed to be controlled by a human rather than by a computer (Bailenson, Blascovich, Beall, & Loomis, 2003). Similarly, Schechtman and Horowitz observed increased interpersonal behavior when the ostensible interactants were human rather than computerized. This effect was found using a visually impoverished task in which equipotent interactants had to cooperate through a text-based interface (2003).

Nonetheless, many if not most human–computer and human–human interactions occur under conditions of *un*equal stature or power (e.g., teacher–student, doctor–patient, supervisor–employee, officer–soldier, interviewer–interviewee), when the evaluations of the superordinate member of the dyad have direct repercussions for the subordinate. Under these conditions, the subordinate must not only engage in the usual social rules, but also must respond directly to the desirable or undesirable feedback the superordinate supplies. This forces the question: Do we respond differently when we are praised (or damned) by a computer rather than another human?

One clue derives from studies of impression management, which show that subordinates display strong interpersonal behaviors toward authority figures (Burgoon, Johnson, & Koch, 1998). Naturally, this literature makes no prediction regarding computerized authorities. We reasoned that, although power asymmetry by itself might not be sufficient to elicit anthropomorphism of the computer, receiving evaluative feedback (whether desirable or undesirable) from that authority may strengthen interpersonal reactions. For instance, if participants feel rejected by the authority figure, perhaps their emotional reaction will be strong enough to provoke social responses toward that authority even when doing so is gratuitous. Due to the scant literature on this topic, this prediction remained speculative.

To gauge the impact of labeling and subordination on SRCT, we asked participants to take part in individual mock job interviews, in which they answered questions that were delivered over an audio link. The interviewer at the other end of the link was stipulated to be either a human speaking over a microphone or a vocalizing computer program. This arrangement allowed determining whether the putative identity of the remote interviewer would affect participants' interpersonal responses during a mediated exercise of asymmetric power. Following the mock interview, the interviewer (whether "computer" or "human") informed the participant whether he/she was accepted for or rejected from the job. This second manipulation was designed to indicate whether evaluative feedback would moderate any effects of the interviewer's alleged identity.

We examined social reactions by measuring nonverbal behavior (particularly facial displays), evaluative remarks about the interviewer, and emotion-report before and after the interview. We used nonverbal, cognitive, and emotional measures because they are well-

established indicators of social influence (Argyle, 1972; Fridlund & Russell, 2006; Kendon, 1981; Lang, 1968; Mandler, Mandler, Kremen, & Sholiton, 1961).

2. Method

2.1. Design

This experiment consisted of two phases. In Phase 1, participants were told they would be interviewed for a mock job by either a *human* or an artificially intelligent *computer*; these constituted the two categories of *interviewer identity* (ID). During the interview, participants' verbal and nonverbal displays were recorded and measured. In Phase 2 of the experiment, the participants from each ID category were informed that they were either accepted for or rejected from the job. This manipulation comprised a 2×2 factorial design, counterbalancing ID with *interviewer feedback* (FB). After receiving feedback, participants' emotion self-reports and impressions of the interviewer were ascertained. All conditions were between subjects with equal numbers of males and females in each of the four conditions, and the order of conditions was counterbalanced across participants.

2.2. Participants

Participants consisted of psychology students from the University of California, Santa Barbara (20 male, 20 female), who received introductory psychology course credit for their participation. Age ranged from 17 to 23 yr ($M = 20$ yr). The entire research protocol was approved by the UC Santa Barbara Committee on Research.

2.3. Measures

In order to assess participants' responses, observations of three types were conducted – nonverbal behavior, self-reported emotion, and evaluative impressions of the interviewer. Nonverbal behaviors were tallied by blind judges, and included judgments about the presence (or absence) of facial displays and other behaviors indicating social responsiveness. Nonverbal behaviors that were assessed included facial displays such as smiles, frowns, and "yuck" faces, and adjunctive behavior such as silence-fillers, self-manipulations (e.g., nose-scratches), and signs of politeness and embarrassment. All nonverbal behaviors were rated using a binary scale ("present" or "absent").[1] (See Appendix A for definitions of these measures.)

Participants' emotional reactions were assessed via self-report questionnaire, administered before and after the experimental manipulation, here referred to as the pre-stimulus and post-stimulus emotion forms. The questionnaire included 7-point Likert-type questions asking the participant how he/she felt at that moment. These items included questions like "To what extent do you feel happy?," and "To what extent do you feel tired?" Additionally, we sought to obtain a general motivational factor (perceived "enthusiasm") without asking their motivation directly (and reactively). So, a composite factor

[1] Eight additional exploratory nonverbal measures were coded by the blind judges which were excluded from analysis on the basis of insufficient power (<.45 for each item). This was not surprising as the measures were exploratory only and unvalidated. Details about these measures available upon request.

was constructed to gauge how "enthusiastic" the participant was before and after the interviewer. This was achieved by combining scores from the items "excited," "curious," and "alert" (tired reverse-coded). Assessing this aspect of the participant's motivation by pooling several related responses should have augmented its reliability, although admittedly it was neither psychometrically constructed nor factorially pure. (See Appendix A for complete list of measures.)

Participants' evaluative impressions of the interviewer were also assessed via self-report questionnaire, administered after receiving feedback about their acceptance for or rejection from the job. This questionnaire, referred to as the impressions form, was designed to gauge on a 7-point Likert-type scale the extent to which participants judged the interviewer to be competent and enjoyable. "To what extent did the interviewer capture your strengths?," and "To what extent do you think the interviewer was sociable?" (See Section 3 for complete list of measures.)

In addition to our three categories of observations, we also collected data about participants' demographic profiles, their level of experience with interviews, and their level of experience with computers. These were obtained on a form labeled the personal information form. Finally, we measured the extent to which participants believed the identity of the interviewer was consistent with the ID condition to which they were assigned. This manipulation check was administered in the form of a questionnaire called the post-interview form.

2.4. Stimuli and apparatus

Interview stimuli included 10 questions, prerecorded by a male human voice, played in a single order to the participant, who was seated in front of the computer keyboard, CRT monitor, and speakers. These questions included topics such as "How did you adjust to your first year of college?" and "What kinds of activities have you been interested in outside of school?" (The interview questions are available upon request.)

Although the interviewer was stipulated to be either human or computer, in actuality the same prerecorded stimulus was used for both – and was altered to be believable in either case. To accomplish this, the voice recording was passed through a preset digital filter that distorted the vocal quality and was pretested for ambiguity as to the source (human or computer) ($n = 23$).[2] All audio editing was implemented using Melodyne 1.5 (Celemony) and Wavelab 4.0 (Steinberg) with the Orange Vocoder plug-in (Prosoniq). Melodyne was used to filter out prosodic variation in the audio clips. Wavelab was then used to apply the preset distortion filter on these clips. (The vocoder specifications are available upon request.)

During the interview, the computer monitor displayed only the appropriate question number (question 1, question 2, etc.) for each interview question. A small lapel microphone detected the participant's voice. Video from a covert camera, as well as the microphone audio, were fed to a VHS video recorder, which taped the participant's entire session. The videotape allowed us to record information about the participant's facial displays, gesturing, and posturing.

[2] Due to a limited sample size, statistical differences between vocal stimuli were non-significant, however means for each level were in the predicted direction.

2.5. Procedure

After reading and signing a general information form, the participant completed the personal information form and the pre-stimulus emotion form at a standard desk space. Following completion of these forms, participants were situated in front of a 17-in desktop computer monitor and asked to complete an online "Prescreening Questionnaire" designed to collect demographic data relevant to a mock job-hiring interview. The questionnaire was used to lend extra face validity to the ensuing interview.

Upon completion of this online questionnaire, participants received standard instructions explaining that they would be asked to answer questions about themselves in a mock interview. The study, they were told, concerned the effectiveness of Internet-based job interviews, which required the interviewer to be elsewhere (ostensibly, a nearby lab).

Depending on the experimental condition, participants were further told that the interviewer would be either a live human being or an artificially intelligent, vocal computer program. Each interview condition was accompanied by a unique justification for the interviewer's voice. When the interviewer was purportedly human, the justification was that the speaker's voice would be filtered in real time to simulate limited-bandwidth Internet transmission. When the interviewer was ostensibly a computer, interviewers were told that the speaker's voice was generated by an advanced computer application that specialized in speech production and analysis.

In both conditions, participants were told that the interviewer alone would generate questions based on their Prescreening Questionnaire, issue the questions over the computer speakers, analyze their responses, and then judge whether they merited being hired for the job by comparing his/her answers to those of the others in the study. They were told that the interviewer followed strict procedures for making this decision, but we could not inform them of the admission criteria, nor the exact nature of the job, until the end of the experiment.

Participants were told that they would be given 45 s to answer each question. Finally, it was explained that all answers were fully confidential and anonymous and that they were not obligated to answer any question that made them uncomfortable.

While the participant was completing the online questionnaire, the experimenter left the room, ostensibly to ready the fictitious "other lab." When the experimenter returned, he/she affixed the microphone and told participants that it would be transmitting their answers to the interviewer. The participant was then left in privacy for the duration of the interview. Following the last interview question, the participant received predetermined feedback telling him/her whether or not he/she got the job. The experimenter then gave the participant the post-stimulus emotion form, impressions form, and post-interview form. After their completion, the participant was fully debriefed. Finally, the experimenter disclosed the covert videotaping and administered a videotape consent form to permit the use and retention of the videotaped session.

3. Results

Response coding and reliability. Audiovisual data were coded by two independent blind judges who were trained to assess the presence/absence of various nonverbal displays. The judges recorded the presence or absence of smiles, frowns, "yuck" faces, silence fillers, self-manipulations, and signs of politeness and embarrassment throughout the interview.

Participants' self-reported emotion ratings before and after the interview, as well as their impressions of the interviewer, were coded using 7-point Likert-type scales.

For each of the seven dichotomous behavioral measures, a single composite score was computed by averaging across the individual scores for the 10 interview questions (i.e., smiles, frowns, yuck faces, silence fillers, self-manipulations, politeness, and embarrassment). All of the composites scores showed non-normal distributions. Three of them (smiles, silence fillers, and embarrassment) were successfully normalized using a logarithmic transformation [$X = \log_{10}(x + 1)$]. The remaining four composite scores could not be logarithmically corrected and were excluded from further analysis.

Interrater reliability (Cronbach's Alpha) was established between the two blind judges. Smiles and silence fillers were significantly correlated, $\alpha = .804$ and $\alpha = .621$, respectively. Embarrassment was not reliable between raters, $\alpha = .061$, so this measure was excluded from further analysis.

Effect of interviewer identity. We first wished to assess if ID had any effect on the two dichotomous variables, smiles and silence fillers. These did differ with ID, $F(1,39) = 5.45$, $p < .05$ and $F(1,39) = 7.94$, $p < .01$, respectively. Specifically, participants were more likely both to smile and to exhibit silence-filling behavior when the interviewer was human (see Figs. 1 and 2; Tables 1 and 2 for means).

Effects of feedback with interviewer identity. Data were also collected on participants' emotion ratings. First, we searched for changes in emotion as a function of ID and FB, controlling for pre-stimulus emotion scores. Thus, we subjected each post-stimulus emotion score to an analysis of co-variance, with the corresponding pre-stimulus emotion score as the covariate. We found no independent effects of ID and no interactions (All F-values < 3.0, $p > .05$). We did, however, find several effects for feedback type. Participants were more curious, excited, and "enthusiastic" after receiving desirable feedback from the interviewer, as compared to undesirable feedback, regardless of interviewer type, $F(1,39) = 4.91$, $p < .05$; $F(1,39) = 16.01$, $p < .0001$; $F(1,39) = 12.366$, $p < .001$, respectively.

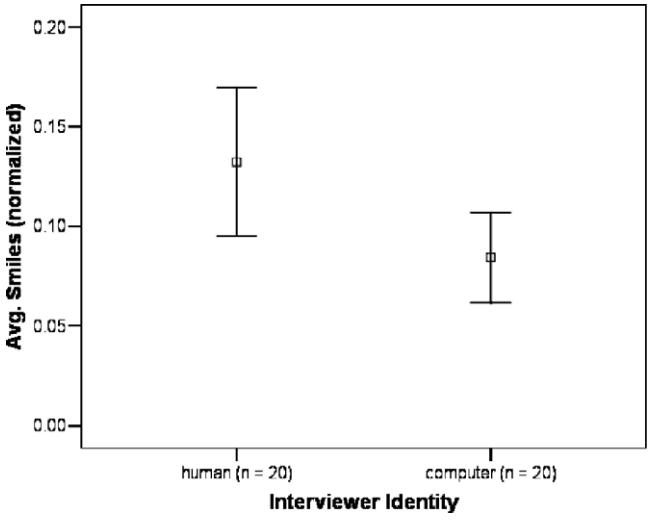

Fig. 1. Average number of smiles toward human and computer interviewers, $F(1,39) = 5.45$, $p < .05$.

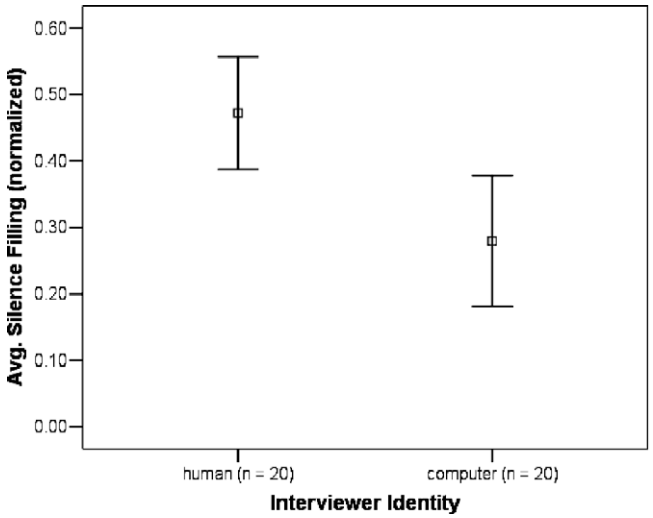

Fig. 2. Silence filling behavior toward human and computer interviewers, $F(1,39) = 7.94$, $p < .01$.

Table 1
Average smiles toward human and computer interviewers (from 0 to 1)

ID	Mean	SD	*N*
Human	.132	.084	20
Computer	.084	.050	20

Table 2
Average silence fillers toward human and computer interviewers (from 0 to 1)

ID	Mean	SD	*N*
Human	.472	.190	20
Computer	.280	.221	20

We also found a marginal effect in which participants tended to be less frustrated after receiving desirable feedback compared to undesirable feedback, $F(1,39) = 3.71$, $p = .062$. No differences were found in reported happiness, sadness, boredom, anxiety, anger, or relaxation (All *F*-values < 2.3, $p > .05$; see Table 3 for means).

Attributions toward interviewer. We then tried to discern whether ID affected participants' impressions of and attributions toward the interviewer. Overall, people rated the interviewer more positively when FB itself was desirable, regardless of ID. Specifically, people rated the interviewer as more *sociable,* $F(1,39) = 6.15$, $p < .05$, and more *likeable,* $F(1,89) = 7.06$, $p < .05$, when they were accepted to the mock job. Other ratings related to likeability (nice, fun, and casual) suffered from weak statistical power, so we pooled them with the above-mentioned factors and found a corresponding effect, $F(1,39) = 6.33$, $p < .03$, independent of ID.

Participants receiving desirable feedback also reported that the interviewer was better able to *capture their strengths,* $F(1,39) = 7.67$, $p < .01$, and that the interviewer enabled them to give a more *accurate impression*, $F(1,39) = 8.60$, $p < .01$. As above, other ratings

Table 3
Emotion ratings pre- and post-interview

Emotion	ID	FB	*M*(pre-)	*M*(post-)	*n*
Curious	Human	Rejection	4.90 (1.370)	3.60 (2.221)	10
		Acceptance	4.90 (1.449)	5.40 (1.838)	10
	Computer	Rejection	5.60 (1.174)	5.00 (1.732)	9
		Acceptance	4.60 (.699)	5.60 (.699)	10
Tired	Human	Rejection	5.00 (1.700)	4.60 (2.221)	10
		Acceptance	4.30 (1.567)	2.80 (1.317)	10
	Computer	Rejection	3.50 (1.650)	2.89 (1.764)	9
		Acceptance	5.40 (1.647)	4.40 (1.838)	10
Happy	Human	Rejection	4.50 (.972)	4.60 (1.350)	10
		Acceptance	4.55 (1.462)	5.10 (1.101)	10
	Computer	Rejection	4.10 (1.287)	4.33 (1.414)	9
		Acceptance	3.90 (.738)	4.90 (.994)	10
Sad	Human	Rejection	1.70 (.949)	2.20 (1.398)	10
		Acceptance	1.70 (.483)	1.50 (.972)	10
	Computer	Rejection	2.60 (1.838)	2.33 (1.225)	9
		Acceptance	2.80 (1.135)	2.60 (1.430)	10
Bored	Human	Rejection	2.90 (1.792)	2.40 (1.647)	10
		Acceptance	2.50 (1.269)	1.80 (.789)	10
	Computer	Rejection	3.30 (1.418)	2.44 (1.130)	9
		Acceptance	2.50 (1.354)	2.10 (.994)	10
Anxious	Human	Rejection	3.10 (1.912)	2.40 (2.066)	10
		Acceptance	3.50 (1.434)	3.10 (2.183)	10
	Computer	Rejection	4.00 (1.333)	3.33 (1.871)	9
		Acceptance	4.10 (1.449)	4.10 (1.853)	10
Excited	Human	Rejection	2.50 (1.269)	1.80 (1.135)	10
		Acceptance	3.80 (1.549)	4.70 (1.767)	10
	Computer	Rejection	3.90 (1.197)	3.33 (1.500)	9
		Acceptance	3.60 (1.350)	4.30 (1.636)	10
Frustrated	Human	Rejection	1.80 (1.476)	2.00 (1.886)	10
		Acceptance	1.70 (.823)	1.30 (.675)	10
	Computer	Rejection	2.00 (1.247)	2.33 (1.414)	9
		Acceptance	3.10 (2.132)	2.40 (1.776)	10
Angry	Human	Rejection	1.60 (1.265)	1.90 (1.912)	10
		Acceptance	1.20 (.632)	1.30 (.675)	10
	Computer	Rejection	1.80 (1.874)	1.33 (.500)	9
		Acceptance	2.00 (1.247)	1.50 (.707)	10
Relaxed	Human	Rejection	5.10 (1.197)	5.20 (1.549)	10
		Acceptance	4.20 (1.687)	4.20 (1.135)	10
	Computer	Rejection	4.90 (1.197)	4.67 (1.500)	9
		Acceptance	4.00 (.943)	4.60 (1.174)	10
"Enthusiastic"	Human	Rejection	3.47 (.789)	2.93 (1.350)	10
		Acceptance	4.13 (1.146)	5.10 (1.089)	10
	Computer	Rejection	4.67 (1.042)	4.48 (1.281)	9
		Acceptance	3.60 (.886)	4.50 (1.189)	10

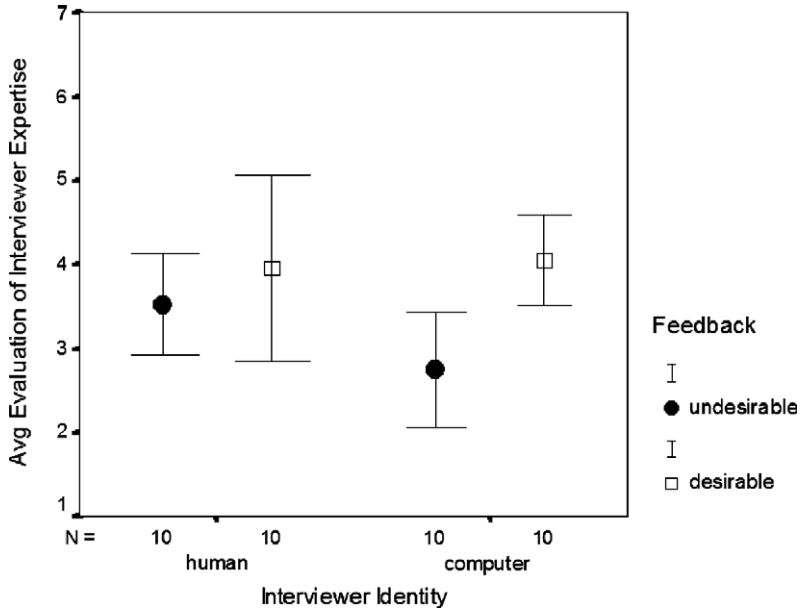

Fig. 3. Evaluations of "Expertise" a function of desirable FB but not ID, $F(1,39) = 6.45$, $p < .05$.

related to expertise were found to have weak statistical power, and so these latter two factors were subsequently pooled with two other factors (competence and objectivity) to create the factor named "expertise." It was found that evaluations of "expertise" also increased as a function of desirable FB but not as a function of ID, $F(1,39) = 6.45$, $p < .05$ (Fig. 3). Participants who received undesirable feedback from the computer interviewer showed marginal tendencies to demote the computer interviewer's intelligence, $F(1,39) = 3.11$, $p = .086$, and objectivity, $F(1,39) = 3.07$, $p = .089$, relative to the other conditions. The other five items did not differ as a function of ID (see Table 4 for means). There is no strong evidence to conclude that the identity of the interviewer had an impact on the above impressions of the interviewer. However, because statistical power is low (<.3), this pattern is vulnerable to Type II error.[3]

Finally, we examined whether the beliefs that the participants reported about the interviewer identity were consistent with the condition to which they were assigned. First, we transformed the belief question in each condition into a single continuous variable from 0 ("don't believe") to 1 ("believe"). We subjected this variable to a one-way analysis of variance, with ID as the independent factor and observed a significant difference in which

[3] A series of three-way ANOVAs was also used to test the effects of ID, FB, and participant sex on impressions of the interviewer. These tests showed no affect of ID; however, one yielded a two way interaction in which female participants who received undesirable feedback rated the interviewer as significantly less able to capture their strengths than did females receiving desirable feedback ($F(2,38) = 4.5$, $p < .05$). We then conducted a series of post hoc tests (Fisher's LSD, Tukey's HSD, & Bonferroni) to observe the interaction while accounting for the inflation of *alpha* that results from three-way ANOVAs. Only Fisher's LSD showed significance, and did so in the predicted direction (All MD > 1.8, $p < .05$), suggesting that the female tendency to demote the interviewer's expertise on the basis of undesirable feedback occurred regardless of ID. But because statistical power was weak (<.1), it remains possible that the interaction was due to chance.

Table 4
Evaluative impressions of human and computer interviewers after acceptance or rejection

Evaluation	ID	FB	*M*	*n*
Competent	Human	Rejection	3.89 (1.364)	9
		Acceptance	3.70 (2.058)	10
	Computer	Rejection	3.40 (1.430)	10
		Acceptance	4.00 (.667)	10
Nice	Human	Rejection	3.22 (1.922)	9
		Acceptance	3.80 (1.687)	10
	Computer	Rejection	3.00 (1.633)	10
		Acceptance	4.20 (1.135)	10
Fun	Human	Rejection	2.00 (1.581)	9
		Acceptance	2.50 (1.958)	10
	Computer	Rejection	2.00 (1.155)	10
		Acceptance	2.70 (1.767)	10
Sociable	Human	Rejection	2.00 (1.500)	9
		Acceptance	2.70 (1.947)	10
	Computer	Rejection	1.70 (.823)	10
		Acceptance	3.40 (1.897)	10
Objective	Human	Rejection	4.33 (1.871)	9
		Acceptance	4.10 (1.370)	10
	Computer	Rejection	3.00 (1.700)	10
		Acceptance	4.50 (1.179)	10
Intelligent	Human	Rejection	4.00 (1.581)	9
		Acceptance	3.20 (2.150)	10
	Computer	Rejection	3.20 (1.989)	10
		Acceptance	4.40 (1.430)	10
Casual	Human	Rejection	3.22 (1.716)	9
		Acceptance	4.20 (2.251)	10
	Computer	Rejection	3.40 (2.171)	10
		Acceptance	4.60 (1.174)	10
Slang	Human	Rejection	4.33 (1.414)	9
		Acceptance	4.00 (2.357)	10
	Computer	Rejection	3.90 (2.283)	10
		Acceptance	3.60 (2.011)	10
Impression	Human	Rejection	3.11 (1.364)	9
		Acceptance	4.40 (1.897)	10
	Computer	Rejection	2.20 (1.687)	10
		Acceptance	4.00 (1.247)	10
Strengths	Human	Rejection	2.11 (.782)	9
		Acceptance	3.60 (2.011)	10
	Computer	Rejection	2.40 (1.174)	10
		Acceptance	3.70 (1.160)	10
Likability	Human	Rejection	2.67 (1.414)	9
		Acceptance	3.30 (1.889)	10
	Computer	Rejection	2.20 (1.229)	10
		Acceptance	4.00 (1.563)	10
"Expertise"	Human	Rejection	3.53 (.837)	9
		Acceptance	3.95 (1.563)	10
	Computer	Rejection	2.75 (.957)	10
		Acceptance	4.05 (.744)	10

people in the "computer" condition reported a stronger belief in the purported identity of the interviewer than did those in the "human" condition, $F = 18.102$, $p < .001$. Cell means illustrate that participants in the "computer" condition did indeed believe in the interviewer's "computer" identity ($M = .90$, $SD = .308$) whereas participants in the "human condition" confessed some suspicion of the interviewer's "human" identity ($M = .35$, $SD = .489$).

4. Conclusion

Do people differ in their social reactions to an interviewer depending upon whether the interviewer is human or a computer? And does the interviewer's judgment matter? Our results reveal a mixed answer. Participants did not report feeling any happier with the human interviewer, but they smiled more toward him. They did not think the human was nicer or more likeable, sociable, or fun, but they made greater efforts to speak to him. They did not consider the human to be more of an expert than the computer, but they were more affected by a rejection from the human.

Participants' reactions to the interview feedback also yielded effects that were independent of the interviewer's identity. Even though all participants were fully aware that the interview was only a simulation, they were emotionally affected by information about their acceptance to or rejection from the job: They were more curious, excited, and "enthusiastic" after being accepted. Other emotion ratings, however, did not differ. Furthermore, likeability ratings and evaluations of expertise were strongly predicted by acceptance versus rejection, particularly for females.

Our mixed results follow a consistent pattern. Nonverbal displays (smiling and silence filling) were consistently responsive to the human. Both self-reported emotion and impressions of interviewer likeability and expertise were consistently responsive to the nature of the interview feedback independent of ID. Thus, the behavioral results complement the findings of Schechtman and Horowitz (2003), and suggest that in a context of asymmetric power, people are more communicative toward ostensibly human interactants than toward computer interactants, even when the behavior of both interviewers is identical (For related findings, see Bailenson et al., 2003; Bonito et al., 1999; Lee & Nass, 2002; Martin & Nagao, 1989; Resnik & Lammers, 1985). This pattern was not confirmed by the emotional and cognitive reactions to the interviewer, which were sensitive to feedback only.

5. Discussion

We consider a limited number of ways to explain the consistent, but mixed pattern of results. First, why did participants smile more to the human interviewer if they did not report feeling any happier having interacted with him? One explanation for this discrepancy is that the smiles emitted in this context are not expressions of emotion so much as communications to the interviewer, much as we commonly make expressions to others on the telephone (Fridlund, 1994, 1997). An analogous argument explains why participants exhibited greater effort speaking to the human despite the fact that they did not consider him to be nicer or more likeable, sociable, or fun, than the computer interviewer. Likeable or not, efforts to fill silences with topic-relevant speech suggest that participants were operating under a norm of reciprocity in their communications with the interviewer. That this pattern was not found with computer interviewers is evidence that computers

were not regarded as independent agents of communication that should call upon such a reciprocity norm.

If nonverbal behavior is more responsive to the human interviewer, why do we not find a similar pattern in participants' reported emotion? Could it be that this equal sensitivity to both interviewers is evidence of a social reaction toward the computer as well as the human? Although this interpretation is plausible, it fails because the absolute values of the means for these emotion ratings converge near the neutral point of the Likert scale. This suggests that, in absolute terms, participants are not particularly happy, sad, excited, etc., after interacting with either interviewer. Although they are clearly emotionally affected by different types of feedback, this is not associated with interviewer identity. Therefore, by observing solely the emotion ratings, we cannot find evidence that either interviewer qualifies as an agent of social exchange. Indeed, if participants were emotionally partial to either interviewer, the fact that this was only a *mock* interview may have been sufficient to motivate them to under-report the magnitude of their emotional response. Such demand characteristics are a common problem in experimental simulations, and future studies of this kind may overcome this problem by using less reactive ways of ascertaining emotion (but see Fridlund & Russell, 2006).

A related question is that if participants' nonverbal behaviors are "human-biased," then why did not they report similar biases in cognitive appraisals of the interviewer? For example, why was not the human deemed more expert than the computer? Because evaluations of the interviewer were relatively anonymous, we see no compelling reason why participants should be motivated to misreport their impressions as they might have with their emotion ratings. Instead, we believe that participants evaluated both interviewers' expertise the same, not because they awarded the computer with humanlike expertise, but because they felt neither interviewer was particularly expert. Again, the absolute values of the means suggest this is the case. Rather than a commentary on social reactions toward the computer, we feel this result signals a ceiling effect for expertise. We suggest that future studies on this topic take precautions to ensure that the stimuli are designed to accommodate to higher ratings of expertise.

One reason why emotional and evaluative impressions of the interviewer were insensitive to identity could be that people reported a moderate suspicion that the "human" interviewer was in fact human. We feel that this is an insufficient explanation for our effects because if participants were truly suspicious of the interviewer's identity, we would not have observed the pronounced difference in allocation of smiles and silence-filling behavior between the two interviewer types. So, even if they were suspicious to some degree, this did not appear to diminish their special responsiveness to the human interviewer. It may also imply that, even when participants were suspicious of the interactant's "human" identity, they classified the interactant as human for the purposes of the interaction.

Generally speaking, our behavioral measures appear to paint a picture that only a human merits social, communicative exchange, and that mere labels about an interactant's identity are sufficient to produce this effect – even when participants may be suspicious of that interactant's identity. Our cognitive and emotional measures supplement this conclusion, indicating that some social influence between human and computer can emerge by receiving trivial feedback about one's acceptance to or rejection from a mock job, independent of interviewer identity.

It has been argued elsewhere that self-report measures by themselves are oftentimes inadequate because participants can be poor judges of their own thoughts and feelings

(Bailenson et al., 2004). Consequently, self-report measures are most appropriate in conjunction with behavioral measures, which can be less obtrusive and more sensitive to more automatic social signals (Bailenson et al., 2004; Cook & Campbell, 1979). Heeding this advice, this experiment may lend its strongest support to the argument that identity labeling alone can be sufficient to elicit rich social responses toward ostensibly human interviewers in a way that is not matched by SRCT. However, evaluative feedback from the interviewer can elicit some rich impressions and emotional reports that do not necessarily discriminate between either identity.

Taken together, our results can be interpreted in terms of impression management, whereby the individual defers to and attempts to engage and appease others because others' evaluations bear social consequences (Baumeister, 1982b; Kacmar & Carlson, 1999). In this case, the audio computer interviewer did not appear to merit the social status achieved by an ostensible human except in regard to some evaluative components of the interaction. It may be that a stimulus labeled "human" can afford to be lacking in a host of other diagnostic cues, such as naturalistic vocal and visual cues, whereas a stimulus labeled "computerized" must advertise much richer audiovisual attributes in order to elicit the same level of social reactions so easily directed toward ostensible humans. This sensitivity change is consistent with established models of social influence (Blascovich, 2002).

Many questions, however, remain unanswered, such as how these interpersonal responses may differ in contexts in which the participant has more power than the interactant, and when communication occurs along more than audio channels. Pursuing these research avenues using standard social psychological methods may be quite informative about various facets of human–computer interaction.

Acknowledgements

We thank Chris McCallister, Sarah Haskell, and Jay Smith for their assistance with data collection, and Daphne Bugental and Nesta Aharoni for editorial comments.

Appendix A.

Emotion form

To what extent are you feeling each of the following right now?

Circle a number between 1 and 7. (1 = not at all, 4 = somewhat, 7 = very much):

Curious	Tired
Happy	Sad
Bored	Anxious
Excited	Frustrated
Angry	Relaxed

Impressions form

Circle a number between 1 and 7. (1 = not at all, 4 = somewhat, 7 = very)

- To what extent did you feel that the interviewer was competent?
- To what extent did the questions enable you to give an accurate impression of yourself?
- To what extent did the questions chosen capture your strengths?

- To what extent did you feel that the interviewer's decision was objective?
- To what extent did you feel that the interviewer was intelligent?
- To what extent did you feel that the interviewer was nice?
- To what extent did you feel that the interviewer was fun?
- To what extent did you feel that the interviewer was sociable?
- To what extent did you like the interviewer?
- To what extent was your relationship with the interviewer casual and informal?
- To what extent did you feel comfortable using slang words in front of the interviewer?

Ratings from blind judges
For each interview question, rate the presence or absence of:

- Smiles – self-explanatory.
- Frowns – self-explanatory.
- "Yuck" faces – displays of disgust.
- Silence fillers – statements produced after a conclusive pause in speech.
- Self-manipulation – e.g., picking nails, twirling hair, biting lips, scratching face.
- Signs of politeness – e.g., saying "thank you" or "your welcome.
- Signs of embarrassment – e.g., blushing, nervous laughter.

References

Argyle, M. (1972). Non-verbal communication in human social interaction. In R. A. Hinde (Ed.), *Nonverbal communication* (pp. 243–269). New York: Cambridge University Press.

Bailenson, J., Aharoni, E., Beall, A., Guadagno, R., Dimov, A., Blascovich, J. (2004). Comparing behavioral and self-report measures of agents' social presence in immersive virtual environments. In *Proceedings of the 7th annual international workshop on PRESENCE* (2004, Valencia, Spain).

Bailenson, J., Blascovich, J., Beall, A., & Loomis, J. (2003). Interpersonal distance in immersive virtual environments. *Personality & Social Psychology Bulletin, 29*(7), 819–833.

Baumeister, R. (1982b). A self-presentational view of social phenomena. *Psychological Bulletin, 91*(1), 3–26.

Blascovich, J. (2002). Social influence within immersive virtual environments. In R. Schroeder (Ed.), *The social life of avatars* (pp. 127–145). Springer-Verlag.

Bonito, J., Burgoon, J., Bengtsson, B. (1999). The role of expectations in human–computer interaction. In *Proceedings of the international ACM SIGGROUP conference on supporting group work*.

Burgoon, J. K., Johnson, M. L., & Koch, P. T. (1998). The nature and measurement of interpersonal dominance. *Communication Monographs, 65*(4), 308–335.

Cook, T. D., & Campbell, D. T. (1979). *Quasi-experimentation: Design & analysis issues for field setting*. Boston, MA: Houghton Mifflin Co.

De Laere, K., Lundgren, D., & Howe, S. (1998). The electronic mirror: human–computer interaction and change in self appraisals. *Computers in Human Behavior, 14*(1), 43–59.

Fridlund, A. J. (1994). *Human facial expression: An evolutionary view*. San Diego, CA: Academic Press.

Fridlund, A. J. (1997). The new ethology of human facial expressions. In J. A. Russell & J. Fernandez-Dols (Eds.), *The psychology of facial expression* (pp. 103–129). Cambridge: Cambridge University Press.

Fridlund, A. J., & Russell, J. A. (2006). The functions of facial expressions: what's in a face? In V. Manusov & M. L. Patterson (Eds.), *The sage handbook of nonverbal communication*. Thousand Oaks, CA: Sage.

Kacmar, K. M., & Carlson, D. (1999). Effectiveness of impression management tactics across human resource situations. *Journal of Applied Social Psychology, 29*(6), 1203–1315.

Kendon, A. (1981). Introduction: current issues in the study of "nonverbal communication. In A. Kendon (Ed.), *Nonverbal communication, interaction, and gesture* (pp. 1–53). Paris: Mouton.

Lang, P. J. (1968). Fear reduction and fear behavior: problems in treating a construct. In J. M. Shlien (Ed.). *Research in psychotherapy* (Vol. 1, pp. 90–103). Washington, DC: American Psychological Association.

Lautenschlager, G., & Flaherty, V. (1990). Computer administration of questions: more desirable or more socially desirable? *Journal of Applied Psychology, 75*(3), 310–314.

Lee, E., & Nass, C. (2002). Experimental tests of normative group influence and representation effects in computer-mediated communication: when interacting via computers differs from interacting with computers. *Human Communication Research, 28*(3), 349–381.

Mandler, G., Mandler, J. M., Kremen, I., & Sholiton, R. D. (1961). The response to threat: relations among verbal and physiological indices. *Psychological Monographs, 75*(9 whole No. 513), 22.

Martin, C., & Nagao, D. (1989). Some effects of computerized interviewing on job applicant responses. *Journal of Applied Psychology, 74*(1), 72–80.

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

Nass, C., Moon, Y., & Carney, P. (1999). Are people polite to computers? Responses to computer-based interviewing systems. *Journal of Applied Social Psychology, 29*(5), 1093–1110.

Nass, C., Moon, Y., Fogg, B., Reeves, B., & Dryer, D. (1995). Can computer personalities be human personalities? *International Journal of Human-Computer Studies, 43*(2), 223–239.

Nass, C., Steuer, J., Henriksen, L., & Dryer, D. (1994). Machines, social attributions, and ethopoeia: performance assessments of computers subsequent to "self-" and "other-" evaluations. *International Journal of Man-Machine Studies, 40*(3), 543–559.

Reeves, B., & Nass, C. (1996). *The media equation: How people treat computers, television, and new media like real people and places*. New York, NY: Cambridge University Press.

Resnik, P., & Lammers, B. (1985). The influence of self-esteem on cognitive responses to machine-like versus human-like computer feedback. *Journal of Social Psychology, 125*(6), 761–769.

Schechtman, N., Horowitz, L. (2003). Media inequality in conversation: how people behave differently when interacting with computers and people. In *Proceedings of the CHI 2003 conference on human–computer interaction*. Ft. Lauderdale, FL, April 5–10.

Parise, S., Kiesler, S., Sproull, L., & Waters, K. (1999). Cooperating with life-like interface agents. *Computers in Human Behavior, 15*(2), 123–142.

Wenger, M. (1991). On the rhetorical contract in human–computer interaction. *Computers in Human Behavior, 7*(4), 245–262.