Computers as Social Actors: Testing the Fairness of Man and Machine

Prabu David, Tingting Lu, Li Cai
Ohio State University

Paper presented to the Fifth Annual International Workshop, Presence 2002, Universidade Fernando Pessoa, Porto, Portugal, October 9-11.

Prabu David is an associate professor in the School of Journalism and Communication at Ohio State University. Tingting Lu and Li Cai are graduate students at the same institution. Address correspondence to pdavid@osu.edu.

Key Words: Anthropomorphic agents, human-computer interaction, social presence, computer as social actors.

Abstract

The results from this study suggest that the addition of simple social cues to a computer environment, such a name, or the photograph of a human can significantly affect judgments of more abstract constructs within the environment, such as the fairness of a quiz. A computer help system without social cues led to higher evaluations of fairness of a quiz compared to an identical help system with these cues. Moreover, when a computer help system without any social cues blamed participants for poor performance on a quiz, ratings of fairness of the quiz were unaffected. But when blamed by a computer help system named Phil, with a small photograph of a white male, ratings of fairness declined significantly among female participants.

1.0 Computers as Social Actors: Testing the Fairness of Man and Machine

Recent research in human computer interaction (HCI) has provided encouraging results on how the computer interface can enhance communication experience. Personified and animated agents have been found to have significant positive effects on learners or users of computer programs (for example, Lester et al, 1997; Okonkwo & Vassileva, 2001). Moreover, Picard (2001) emphasizes the importance of affect and advocates an affective computer interface that responds to a person's emotions. The underlying argument is that humans are essentially social beings and to maximize the benefits of the human-computer interaction, it is important for the computer to emulate some of the social norms and edicts of human communication.

The emphasis on the social and anthropomorphic aspects of the computer, however, is not without its critics. Some researchers caution that by adding extraneous human-like features to the interface, the human-computer interaction can become tedious and ineffective (Walker, Sproull, Subramani, 1994; Kiesler, Sproull, Waters, 1996).

But some of the more intriguing implications of the social aspects of human-computer interaction come from the media equation research conducted by Reeves and Nass (1996) based on a CASA (computers as social actors) framework (Nass & Steuer, 1993). Nass and his colleagues have conducted a series of studies that demonstrate that people routinely respond to computers and other media technologies using social rules practiced in human communication. In general, this work is centered on the premise that by altering just a few characteristics of the computer interface, humans can be "tricked" into responding to computers as they would to other humans. In the original study, Nass and Steuer (1993) found that by simply changing the voice of the recording in a computer, respondents rated computers as autonomous sources and applied social rules such as differentiation between "self" and "other."

Subsequent results from this program of research suggest that humans adopt politeness rules when interacting with computers (Nass, Moon, & Carney, 1999), are susceptible to computer flattery (Fogg & Nass, 1997), can develop team membership with computers (Nass, Fogg, & Moon, 1996), apply gender stereotypes to computers (Nass, Moon, & Green, 1997), and that a minimal set of cues is sufficient to create computer personalities to which people will respond in the same way they would to similar human personalities (Nass, Moon, Fogg, Reeves & Dryer, 1995).

In many of these studies, through careful manipulation checks, Nass and his colleagues have ruled out the possibility that these responses are due to users' mistaken beliefs that computers are human-like, or that computers act as proxies for human programmers.

Taking into account the converging evidence, Reeves and Nass (1996) make the powerful argument that human-media interaction is not radically different from human-human interaction. This argument, drawn from evolutionary psychology, is based on the premise that our brain is hard-wired to communicate according to certain rules. This hard-wiring has evolved over many years and is well suited for human-human communication. When media, a relatively recent invention, are introduced, human-media communication tends to follow the same patterns of human-human communication. Hence, according to Reeves and Nass, the robustness of the findings of the computers-as-social-actors (CASA) framework is simply a manifestation of the old brain trying to cope with new technologies.

2.0 Extending the "Computers as Social Actors" Framework

Drawing from findings in previous research, in this paper we extend the CASA framework to an online quiz environment. As online learning environments continue to grow, one of the popular trends is this area is the use of computer agents. For example, many "edutainment" computer games targeted toward children use animated agents. Also, many online help systems use anthropomorphic help agents with a name and a visual representation. More sophisticated renderings, such as Microsoft's help agent, are animated and display a richer set of social cues.

In the typical CASA study, the researcher picks out a basic theory or finding in social interaction. Then one of the two parties involved in the interaction is substituted with a computer. When the study is run, the findings from human-computer interaction closely parallel the findings that one would predict in human-human interaction.

To develop the framework for an online quiz environment, we began with two interfaces: one rich in social cues and the other poor in social cues. On the basis of earlier work by Nass and Reeves, we predicted that for the interface rich in social cues, human-computer interaction would closely approximate human-human interaction. When the interface is stripped of its social cues, however, computers would seem more inanimate and machine-like and expectations might be lowered.

To tease out this difference between human and machine, we chose fairness as the dependent variable. Generally, fairness or the lack of fairness is strongly associated with humans, their motives, and their subjectivity. Particularly, if a quiz were to be administered via a computer rather than a human, it is more likely to be perceived as fair and objective. In this context, the absence of social cues could be an advantage because it would increase perceptions of fairness.

Also, the absence of social cues could place less demands on the personality aspects of the online help system. An interface stripped of social cues could get away with poor social skills. On the other hand, an interface endowed with social cues would be expected to closely adhere to norms of social etiquette, and deviations from these norms would be judged harshly by the user.

By manipulating social cues in a computer interface of an online test, we tested differences in perceptions of fairness of an online quiz as a function of three social attributes of the online help agent: anthropomorphism, intelligence, and personality.

3.0 Method

3.1 Task

The task consisted of answering a 5-item multiple-choice quiz on ancient history. A website was designed to administer the quiz. The participants, who were students in an introductory class on the History of Human Communication, were assigned to one of two help systems – Phil’s help system, or Computer help system. To answer a question, the users clicked on a hyperlink, which brought up the help system. In other words, the same clues were consistently forced upon the participants before they could answer a question.

3.2 Stimuli

The Phil condition and the computer condition were identical in all respects, except for two anthropomorphic cues: (1) while one was named “Phil’s Help System,” the computer help system was given a generic name “Computer Help System;” (2) a small passport size photograph of Phil was used in the computer condition, which came up whenever help was sought, whereas no photograph was used with in the computer help system.

Based on these two minor differences between the computer condition and the anthropomorphic agent condition, we had hoped to create a difference in perceived social presence, which in turn would lead to differences in perceived fairness of the quiz.

In addition to the anthropomorphic cues, intelligence and personality of the help system also were manipulated. For the purposes of this study, the operational definition of intelligence was agent’s capability to provide clues that result in success on the quiz. The capable help system led to a strong performance on the quiz, with four correct responses out of the five questions, whereas the incapable system led to a weak performance on the quiz, with only one correct response. It is important to note that the clues provided by the capable and the incapable help systems were identical. The clues amounted to reducing the five multiple-choice options to two. At that point, the respondent had to choose between one of two equally plausible answers. The questions and answers were considerably vague with latitude for different interpretations.

Finally, the personality of the agent was defined in terms of modesty and graciousness. While the positive personality was modest and gracious, the negative personality was a braggart and a blamer. The personality manipulation of the help system was presented at end of the quiz after the scores were posted. Immediately after the scores were presented, the help system disclosed its positive or negative personality profile.

Four personality profiles were created for the 2 (capable, incapable) x 2 (positive personality, negative personality) conditions. In the capable/positive condition, the help system was modest and self-deprecating and passed on the credit to the participant. In the capable/negative-personality condition, the help system took all the credit for the successful performance and blamed the participant for the one wrong answer. In the incapable/positive condition, the help system took responsibility for the poor performance, but expressed encouragement. In the incapable/negative condition, the agent blamed the participant for the bad performance.

3.3 Measures

Before exposure to the stimulus, participants provided data on a number of covariates, which included the short-version of the Marlow-Crowne social desirability scale, extent of computer use, familiarity with animated computer agents, self-assessments of knowledge of ancient history, problem solving ability, and some demographics.

The key dependent variable was the perceived fairness of the quiz, which was tapped on a 7-point scale, “1= Strongly disagree,” and “7 = Strongly agree,” that was used to rate the statement “The quiz was fair.”

Intelligence of the help system was evaluated with two questions. One was a 7-point “Strongly disagree” to “Strongly agree” rating of the statement “Online help seemed to be knowledgeable.” The other was a rating on a 7-point semantic differential scale with “smart” and “dumb” as the anchors.

The personality manipulation of the help system was evaluated with a set of 7-point semantic differential items on a variety of attributes including insensitive/sensitive, cold/warm, impersonal/personal, and unfriendly/friendly.

Two items were used to directly assess whether the social cues of the anthropomorphic agent were associated with humans behind the computer interface. These questions were asked toward the end of the experiment after the respondents had rated the quiz and the help system on the other items mentioned in this section. Using a 7-point strongly disagree/strongly agree scale, participants rated the following statements: “I rated the help system as I would rate a real-life tutor,” and “When evaluating the help system in this questionnaire, I thought of the programmer who set up the help system."

3.4 Design and Participants

Participants were recruited from a large introductory communication class on the history of human communication. A total of 322 undergraduate students participated in the study in return for extra course credit. A 2 (Agent Type: Phil's help system, Computer help system) x 2 (Agent's Capability: capable, incapable) x 2 (Agent Personality: positive, negative) between-subjects design was used.

3.5 Procedure

The experiment was set up as an online website and participants accessed the study via the Internet. Once logged in, the participant was randomly assigned to one of the eight between-subject conditions. After filling out a consent form, the covariates were presented, which was followed by the online quiz.

Then post-test ratings were obtained on the fairness of the quiz, attributes of the agent, and some manipulation checks in the order in which they are presented in the measures section. After all the ratings were obtained, students were debriefed.

4.0 Results

The data were analyzed using a 2 (Agent Type) x 2 (Agent's Capability) x 2 (Agent's Personality) between-subjects analysis of covariance, with social desirability introduced as a covariate. This model was used to test each of the variables discussed in this section.

The key dependent variable was the rating of the fairness of the quiz. Manipulation check variables included friendliness, knowledge level and smartness of the help system, and whether the respondent thought of a real-life tutor while evaluating the help system. The social desirability score was calculated by summing the responses to the 13 true/false items, some of which were awarded a point for a true response and others awarded a point for a false response, as indicated by the authors of the scale.

4.1 Manipulation Checks. Friendliness of the agent, agent's capability, and the perception of the help system as a real-life tutor were used to check the effectiveness of the three experimental manipulations, namely agent's personality, agent's intelligence, and the effectiveness of social cues, respectively. See Table 1 for a summary of means.

For *friendliness*, main effect for personality of the help system, F (1, 312) = 8.89, $p < .01$, MSe = 2.03, and capability of the help system, F (1, 312) = 5.81, $p < .05$, MSe = 2.03, were significant. Not surprisingly, a help system with a positive personality was considered friendlier than a help system with a negative personality. Moreover, a system that led to success on four out of five attempts was considered to be friendlier than a help system that led to success on only one out of five tries. None of the interactions were significant.

Next, the extent to which the help system seemed *knowledgeable* was entered as the dependent variable. As intended, the help system that led to success on four out of five questions was perceived as more *knowledgeable* than the help system that led to failure on four out the five questions. Main effect for the success rate of the help system was significant, F (1, 313) = 108.1, p < .001, MSe = 2.49. When ratings on the smart/dumb semantic differential scale were entered as the dependent variable, again only the main effect of success rate of the help system was significant. These findings suggest that the capable help system that led to success was seen as knowledgeable and smart, whereas the help system that led to failure was seen as less knowledgeable.

To test the possibility that the social cues transformed the computer into a surrogate for humans who designed the computer interface, we examined the extent to which respondents reported to have thought of real-life tutor or the computer programmer who set up the system. Both of these measures were analyzed separately. When "thought of real-life tutor" was introduced as the dependent variable, the main effect for the system's capability was significant, F (1, 312) = 21.6, p < .001, MS e = 2.83. When "thought of programmer who set up the help system" was entered as the dependent variable, none of the main effects or interactions was significant. In summary, the expected main effect for agent type (Phil vs. Computer) was not significant for either measure. Although this finding does not provide evidence that people use the agent as a proxy for a real-world referent, encouraged by the significant effects of the other two experimental manipulations, we proceeded to analyze fairness, the key dependent variable.

4.2 Fairness of the Quiz. When the ratings of the fairness of the quiz were entered as the dependent variable and the data submitted to a 2 x 2 x 2 analysis of covariance, with social desirability as the covariate, the main effects for Agent Type, F (1, 313) = 5.50, p < .05, MSe = 2.82, and Capability, F (1, 313) = 98.60, p < .001, MSe = 2.82, were significant. None of the other main effects or interactions were significant. The significant effect of capability is quite predictable. A quiz on which a higher degree of success was achieved was rated as a fairer quiz than one in which students did not achieve a high degree of success. The more critical finding was the main effect for the Agent type, which has interesting implications.

5.0 Discussion

The means presented in Table 1 suggest that a quiz that was aided by a computer help system without social cues was rated to be fairer than a quiz aided by a computer help system with social cues. The addition of a couple of social cues titled the balance in terms of fairness. From the manipulation checks, there is no evidence that these social cues were associated with plausible real-world referents such as a tutor or a computer programmer. Yet, the presence of two simple social cues, a name and a small photographic representation of the face, were sufficient to alter perceived fairness. Another explanation is that the absence of social cues in the computer condition was perhaps seen as a fairer and more objective testing environment.

Although the effect of social cues or anthropomorphic features on fairness supports the general trend of findings from the research on computers as social actors, the absence of a any referential linking between social cues and real-world referents was puzzling. Nass and Steuer (1993) also report a similar pattern of a significant social reaction to a computer in the absence of connection to any real-world social actors. Unfortunately, the absence of association with other real-world referents doesn't allow room for an explanatory mechanism, such as the tacit role of social presence that we had posited.

Hence, we decided to pursue an additional analysis by introducing a fourth factor, namely gender. If the social cues in the Phil condition were successful in generating a social response leading to different perceptions of fairness between man and machine, it is odd that the personality of the help system, which had a significant effect on ratings of friendliness, did not have an effect on fairness of the quiz. The help system with a negative personality blamed the participant for failure to do well on the test. The system with a positive personality was humble and gracious in crediting the respondent for the success. These differences in personality should have interacted with social cues. In other words, when social cues are high, personality attributes should be taken into account in judgments. In the generic computer system, however, the personality attributes would matter less because the user is dealing with a machine, with minimal social intelligence.

While thinking about the possibility of the social interaction with a computer agent, it seemed possible that the gender of the respondent could interact with the gender of the agent. Particularly among women, a male agent with a propensity to blame the user is less likely to be tolerated than a computer with the same offensive personality trait. Hence, in the next step, data were analyzed again with gender of the respondent as an added factor.

5.1 Moderating Role of Gender on Evaluation of Agent's Personality

The data were analyzed using a 2 (Agent Type: Phil, computer) x 2 (Agent's Capability: capable, incapable) x 2 (Agent's Personality: positive, negative) x 2 (Gender: male, female) between-subjects design, with social desirability entered as a covariate. The means are summarized in Table 2.

5.2 Fairness of the Quiz by Gender. When fairness was entered as the dependent variable, main effects for Agent Type, F (1, 305) = 4.04, p < .05, MSe = 2.78, and Agent's Capability, F (1, 305) = 100.72, p < .001, MSe = 2.78, were significant. Essentially, these main effects replicate the findings from the earlier analysis based on a three-factor design. The notable findings with the four-factor design, however, were the two significant interactions. The Gender x Agent Personality, F (1, 305) = 4.04, p < .05, MSe = 2.78, and the Gender x Agent Type x Agent Personality, F (1, 305) = 4.17, p < .05, MSe = 2.78, interactions were significant.

From Table 2, it appears that the significant three-way interaction is characterized by distinct patterns contingent on the gender of the respondent. Among women, although the personality of the agent had a significant impact when the help system was endowed with social cues (positive personality M = 4.4, negative personality M = 3.6), this difference was not apparent when the system was stripped of these cues. Among men, the personality of the help system did not affect the fairness ratings in either condition.

5.3 Reevaluating Manipulations Check Variables. Each of the manipulation check variables was entered individually as a dependent variable and analyzed with a four-factor design and a covariate, in exactly the same way that fairness of the quiz was analyzed.

When friendliness ratings were entered as the dependent variable, only the main effects for Agent Type, F (1, 305) = 4.04, p < .05, MSe = 2.78, and Agent Capability, F (1, 305) = 100.72, p < .001, MSe = 2.78, were significant. None of the other main effects or interactions were significant.

When the ratings of the agent's knowledge were entered as the dependent variable, only the intended main effect for Agent's capability, which led to success or failure on the quiz, was significant, F (1, 304) = 104.79, p < .001, MSe = 2.50.

To determine whether the social cues were interpreted as a proxy for a real-life tutor or a programmer of the system, the judgments obtained for these two ratings were analyzed. For "real-life tutor" none of the main effects or interactions were significant. However, when "thought of a programmer" was entered, two interactions, Gender x Agent Type, F (1, 304) = 15.17, p < .05, MS e = 3.40, and the Gender x Agent Capability, F (1, 304) = 6.5, p < .01, MSe = 3.40, were significant. None of the other effects were significant.

6.0 Discussion and Conclusions

The results from this study suggest that the addition of simple social cues to a computer environment, such a name, or the photograph of a human can significantly affect judgments of more abstract constructs within the environment, such as the fairness of a quiz. A computer help system without social cues led to higher evaluations of fairness of a quiz compared to an identical help system with these cues. This finding is perhaps an endorsement of the notion that an objective machine is fairer than a subjective human.

Also, in a computer environment augmented with social cues the computer is held to higher standards of social interaction norms. It was interesting to find that when blamed by a computer for poor performance on a quiz, ratings of fairness of the quiz were unaffected. But when blamed by a computer help system named Phil, with a small photograph of a white male, ratings of fairness declined significantly.

These robust effects of social cues in a computer environment are not surprising, given prior work by Reeves and Nass. Our findings replicate many of their findings, such the adoption of social interaction rules when social cues are introduced to a computer environment, sensitivity to gender differences when gender is assigned to a computer, and sensitivity to personality attributes when computers are cast as social actors.

Perhaps the most interesting finding is the significant effect of the personality of the agent on the computer programmer who designed the system. In previous studies, no significant associations with real world referents have been found. In this study, when the human aspects of the help system were emphasized with social cues, such as gender and personality, female participants reported that they thought of the programmer, although the same social cues did not have a significant effect on men. With these findings, one could speculate that social cues in a computer environment evoke social presence, which leads to expectations of adherence to social protocols. The role of social presence as an explanatory mechanism, however, was not addressed in this study and is topic for future research.

Table 1. Fairness of the Quiz and Related Variables as a function of Agent’s Social Cues, Capability and Personality

	Agent				Computer			
	Capable		Incapable		Capable		Incapable	
	+ ve	- ve	+ ve	- ve	+ ve	- ve	+ ve	- ve
	n=33	n=43	n=45	n=40	n=31	n=46	n=50	n=34
Fairness of Quiz	5.0	4.9	3.2	2.8	5.5	5.3	3.4	3.4
Manipulation Checks								
Friendly	5.8	5.3	5.5	4.7	5.7	5.1	5.1	4.9
Knowledgeable	6.0	6.0	4.2	4.1	6.3	6.1	4.1	4.3
Thought of real-life tutor	4.9	4.9	4.0	3.9	4.2	4.9	3.6	3.7

Note. + ve and – ve stand for the personality assigned to the computer and the agent.

Table 2. Fairness of the Quiz and Related Variables as a function of Agent's Social Cues, and Personality, and Respondent's Gender

	Agent				Computer			
	Males		Females		Males		Females	
	+ ve	- ve	+ ve	- ve	+ ve	- ve	+ ve	- ve
	n=33	n=43	n=45	n=40	n=31	n=46	n=50	n=34
Fairness of Quiz	3.2	3.4	4.4	3.6	4.4	4.0	4.2	4.0
Manipulation Checks								
Friendly	5.5	4.7	4.9	5.1	5.0	4.9	5.6	5.1
Knowledgeable	5.0	4.4	4.8	5.0	4.7	5.0	5.1	4.7
Thought of real-life tutor	4.6	4.3	4.2	4.2	3.9	4.2	3.8	3.9
Thought of programmer	2.8	3.0	3.5	3.5	3.4	3.4	2.9	3.3

Note. + ve and – ve stand for the personality assigned to the computer and the agent.

References

Fogg, B.J., & Nass, C. (1997). Silicon Sycophants: The Effects of Computers That Flatter. International Journal of Human-Computer Studies, 46, 551-561.

Kiesler, S. Sproull, L., and Waters, K. (1996). A Prisoner's Dilemma: Experiment on Cooperation with People and Human-Like Computers. J. of Personality and Social Psychology, 70(1), 47-65.

Lester, J., Converse, S., Kahler, S., Barlow, S., Stone, B., &Bhogal, R. (1997). The persona effect: Affective impact of animated pedagogical agents. In proceedings of CHI 97, 359-366, ACM Press.

Nass, C. & Steuer, J. (1993). Voices, boxes, and sources of messages: Computers and social actors. Human Communication Research, 19(4), 504-527.

Nass, C., Fogg, B. J., & Moon, Y. (1996). Can computers be teammates? International Journal of Human-Computer Studies, 45(6).

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.

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

Nass, C., Moon, Y., & Green, N. (1997). Are Machines Gender Neutral? Gender-Stereotypic Responses to Computers With Voices. Journal of Applied Social Psychology, 27(10), 864-876.

Okonkwo, C., Vassileva, J. (2001) Affective Pedagogical Agents and User Persuasion, in C. Stephanidis (ed.) Proc. "Universal Access in Human - Computer Interaction (UAHCI)", held jointly with the 9th International Conference on Human-Computer Interaction, New Orleans, USA, 397-401.

Picard, R. & Klein, J. (2001). Computers that Recognize and Respond to User Emotion: Theoretical and Practical Implications, to appear in Interacting with Computers.

Walker, J. H., Sproull, L. and Subramani, R. (1994). Using a Human Face in an Interface. CHI '94 Human Factors in Computing Systems, Boston MA, April 1994, 85-91.