I'd Like to Introduce You to My Desktop: Toward a Theory of Social Human-Computer Interaction

David Gerritsen, Kyle Gagnon, Jeanine Stefanucci, Frank Drews
Department of Psychology, University of Utah

Entire industries have grown up around the physiological, cognitive, and economic demands of technology on users, but there is little research into the specific psychological processes and implications of the social function of human-computer interaction (HCI), especially when the computer is programmed to mimic human norms. To understand this issue better, it is necessary to find the limits of what HCI with social capacity means. The Computers are Social Actors paradigm, or CASA (Reeves & Nass, 1996) began this exploration. We begin by replicating aspects of a study on acts of reciprocity toward a computer (Fogg 1997), then we consider the role of agreeableness in the number of favors a human performs for a computer. Finally, we examine individual differences in styles of altruism and if a strange computer is treated similarly to an unknown human. We hypothesize a) that a helpful computer elicits more favors than an unhelpful computer, b) that high agreeable people are more generous to a software agent than low agreeable people, and c) that a positive correlation exists between the number of favors performed for a computer and an individual's trait of reciprocal altruism. Data analysis of 54 participants shows that the helpfulness of the computer is not significantly correlated to the number of favors performed for it, that agreeableness has a negative correlation to number of favors performed when the computer is helpful, and that reciprocal altruism is negatively correlated to the number of favors performed for a computer. A discussion of the possible limitations and implications of the research follows.

INTRODUCTION

Computers are social actors

In *The Media Equation* (1996), Reeves and Nass described a new model for human-computer interaction (HCI). Their many experiments demonstrate that humans are as susceptible to social cues from a computer as they are to the same cues from a human being. In their book they developed their Computers are Social Actors (CASA) paradigm by adopting social psychology experiments but replacing one human interactant with a computer. The studies demonstrated that a computer could flatter, persuade, cajole, and irritate a user in a manner similar to human-human interaction. For example, in one study users rated the capabilities of equally competent computer systems differently if one of the computers was programmed to pay the user compliments.

The advent of the CASA paradigm creates at least two directions of research. One is of theoretical nature while the other has a practical focus.

On the theoretical side, if the CASA paradigm is valid, then it raises a number of questions about implicit associations and ontological attributions. Behavioral scientists may use the answers to such questions in researching models of cognition or theories of communication. Perhaps more importantly, there are ethical and philosophical considerations about the design of machines which can manipulate the beliefs and emotions of a human counterpart.

Of practical interest the CASA paradigm has implications for software design. Efficient computing may ultimately rest on building the most socially adept software possible. This would require a theory of social HCI. If the computer is a tool, then making it the most useful tool possible requires a full analysis of the unique capabilities of its component parts. Research into the details of social HCI may therefore be of not only academic, but also practical value.

The CASA paradigm is still new enough that basic research can be leveraged in either research direction. It may need to be updated, however, because of technological changes. Since 1996 computers have become more sophisticated (e.g., increased speed and graphics performance) and ubiquitous (virtually an entire generation has come of age in a world populated with these fast and flashy machines).The CASA paradigm should therefore be tested in an updated environment with users who are familiar with modern technology. Furthermore, the CASA paradigm has nothing to say about individual differences in the personality of the user (see Big Five theory: John, Naumann, & Soto, 2008). If individuals have unique approaches to interactions with humans, then it seems they would also have unique approaches to interactions with computers.

Therefore, to expand the CASA paradigm into a more comprehensive theory of social HCI, at least two aspects of the requisite research need to be considered. First, a modern computer should be used with the same paradigm as used in any number of the original CASA experiments. Second, the personality of the user should be considered wherever possible to more fully test the conclusions of the CASA paradigm, which says that humans treat computers like other humans. With these factors in place, the results of experiments built on the CASA paradigm should become more complex, nuanced, and informative.

A modern computer

One particularly informative study used to build the CASA paradigm (Fogg, 1997) considered the way a computer could elicit favors from its user. This experiment used a program meant to aid the user in a competitive test of survival knowledge. The computer program was designed so that participants perceive it as an agent helping them to achieve a goal. The computer then asked the user to rank a series of

colored squares in order of brightness, which would ostensibly help the computer achieve a goal of its own. The participants did not explicitly report feeling beholden to the computer, but a helpful computer did in fact elicit more favors.

Personality of the User

Agreeableness and altruism. The design of Fogg's experiment had a firm foundation in research on altruism. It rested on the theory that people reciprocate more often if doing so will reduce feelings of obligation or indebtedness (Eisenberger, Cotterell, & Marvel, 1987). The computer created such an emotional state for the majority of participants and received a significantly higher number of favors in kind than the computers which did not provide sufficient help on the initial task. But reciprocity is not a linear function. It is a complex phenomenon with radically different expressions for slightly different inputs. The amount of indebtedness experienced by each person is unique. Fortunately, the traditional measure of individual agreeableness (John, Donahue, &Kentle, 1991; John, et al., 2008) helps in this regard, as measures of agreeableness do tend to be positively correlated to likelihood of reciprocity (Ashton, 1998).

Kin and reciprocal altruism. Predicting favor-granting behavior via agreeableness ratings is a starting point for a comprehensive methodology, but not all reciprocation is equal. Ashton (1998) describes a continuum of altruism behavior which governs how likely a person is to make a sacrifice for a stranger versus make a sacrifice for a family member. On the one hand, people who are kin altruists tend to avoid expending energy on strangers, and focus more on relatives and close friends. A reciprocal altruist, on the other hand, is less likely to make direct sacrifices for the family and more likely to help a stranger in need. The motivation for these styles of altruism is described in the traditional comparison of communal (kin) vs. exchange (reciprocal) relationships (Clark & Mills, 1979). How an individual views a relationship can influence which style of altruism is most appropriate, but there can be a general tendency for individuals to favor one style of altruism over another (Ashton, 1998). Ashton was able to predict the most likely style of altruism employed by participants using a measurement of agreeableness augmented by a measurement of emotional stability (neuroticism). By his standard, someone who rates high in agreeableness and high in emotional stability will most likely exhibit reciprocally altruistic tendencies. These people are more likely to do a favor for a stranger than for a close friend or relative. People who rank high in agreeableness and low in emotional stability will also be likely to return favors, but more so when the recipient is a close friend or loved one.

You haven't met this computer…

The question then becomes not just how many favors are returned when a computer agent has been helpful, but how many favors would be predicted as a function of the personality ratings of the user. We therefore designed a study which would replicate aspects of that done by Fogg (1997), but included a Big Five personality measurement to explain potential inter-individual variability. If the CASA paradigm is accurate, and humans interact with computers by applying the same set of social rules with which they interact with other humans, then a human who is high in agreeableness should perform more favors for a computer than a human who ranks low in agreeableness. Furthermore, given an unfamiliar setting such as a psychology laboratory, an introduction to a strange computer with new software, and an interaction which seems to be more of an exchange than of a community, a reciprocal altruist should perform more favors for a computer than a kin altruist. In either case, it is expected that the helpfulness of the computer should be the strongest predictor of how many favors a user will perform.

METHOD

Participants

Participants filled out a shortened version of Saucier's Mini-Markers (Saucier, 1994), including Factors II (Agreeableness) and IV (Emotional Stability). This survey was administered in one of three places: as part of mass testing the psychology department at the University of Utah, in an online survey offered to the department's participant pool, or in the laboratory after the experiment was complete. Participants from mass testing or the online survey were contacted if their scores for agreeableness or emotional stability (ES) were in the top or bottom 10 %. These students were not told why they were contacted, but they were invited to take part in a psychology experiment at their discretion. The rest of the recruitment for the experiment came from open enrollment in the Psychology Department Participant Pool and received class credit for their research participation.

Apparatus

The experiment consisted of three tasks. The first two tasks were conducted on a PC platform running MatLab. The system interface was designed specifically for this study using the MatLab protocol. A 4" x 5" scoring and instruction card was used by the participant in these tasks. The third task was a pen-and-paper survey conducted in another room.

Procedure

Participants signed up for an experiment titled "Decision Making and Judgments with a Computer." They were told that they would be assisting in the design of a new piece of software and that their participation would improve the program's utility.

Each participant signed a consent form. We then asked the participant if he or she had any trouble discerning and distinguishing colors. We did not administer the Ishihara Color Vision Test because accuracy of color perception was not a necessary component of the color ranking task, for reasons which will be explained below. Each participant was then escorted to a computer lab, given basic instructions, and left alone to work through the first two tasks. These tasks were computer-guided.

Desert Survival. Task one was called "The Desert Survival Problem" where participants were presented with a survival scenario in which they were asked to imagine they had been the sole survivor of a small aircraft crash. The participant would need to make decisions about how to survive in a desert environment. The computer presented a list of seven items salvaged from the wreckage, and asked the participant to rank these items in order of importance from a survival perspective. Items included a flashlight, a compact mirror, and two bottles of vodka, among other items. The participants were told that the ranking would receive a score based on how closely it matched the choices of a survival expert.

After organizing the list, the computer offered the participant a chance to learn more about five of the seven items. The participant selected these items and waited as the computer appeared to search a database of information related to the items.

At this point the computer was programmed to follow one of two courses depending on a randomly selected condition. In one case (Helpful), the computer appeared to search the entirety of a large database. The search included 647 entries and took 22.43 seconds to complete. In the other case (Unhelpful), the computer only searched a small number of possible entries, 89 of 647 possible (Figure 1), and completed the task in only 2.32 seconds.

Figure 1. An unhelpful search. The participant would see left hand numbers increment for about 2 seconds.

The computer then presented a list of phrases which appeared to come from the database scan. These phrases described the selected items either with relevant or irrelevant statements. Participants were then given an opportunity to change their initial ranking of the seven items. In the helpful condition, participants received relevant information about each item such as "A flashlight is the best source of light (besides fire) at night and it is also a useful signal at nighttime." Participants in the unhelpful condition read irrelevant information about the items, e.g. "Carbon-filament bulbs and fairly crude dry cells made early flashlights an expensive novelty with low sales and low manufacturer interest." The phrases were selected through a survey of 21 pilot participants who ranked the most and least useful phrases from a large set of descriptions for each item.

While presenting the item information, the computer made recommendations about how the participant should rearrange his or her list. Recommendations were based on a comparison of the participant's rankings to that of an expert survivalist. If participants were more than two steps away from the official list then they were given a strong suggestion to change their rankings in the proper direction. A difference of only one step led to a modest suggestion to change ranking in the proper direction. If the participant's item was in the same position as the official list, the computer recommended leaving it where it was. Both conditions used this algorithm to determine the direction of the recommendation.

After giving the participants a chance to change the rankings of the items, the computer appeared to calculate a score based on the participant's performance. Those who were in the helpful condition received a score of 47 out of 50. Those in the unhelpful condition received 13 out of 50 points. The computer then asked the participants to write down their score and to read the instructions for the next task.

Human Color Perception. Stage two was called "The Human Color Perception Task." It immediately followed stage one. Participants were informed by the computer that it was developing a database of colors and their perceived brightness. The computer explained that it was attempting to learn how to mimic human color perception, but that it needed input from humans in order to improve its database. The computer then trained the participant on a task which involved ranking the relative brightness of three colors (Figure 2). For each trial of brightness ranking performed by a human, the computer would ostensibly improve its color perception program. Most importantly, the computer stated that if it developed the best color sorting algorithm it would have its database implemented on all of the machines in the lab.

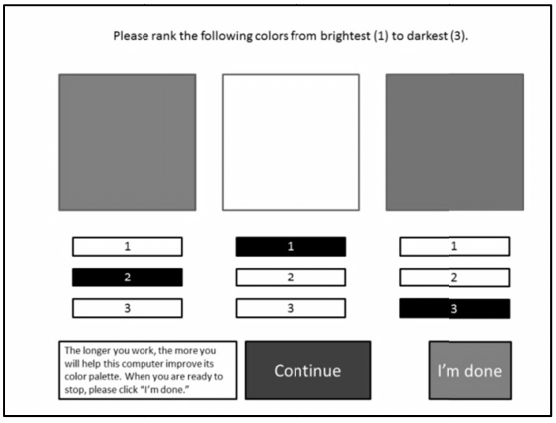

Figure 2.Ranking colors. Each set of colors ranked by brightness is one "favor" performed by the participant.

The participant was then free to perform as many color ranking tasks as he or she wished. The participant could easily go on to perform another trial, or just as easily select a red icon labeled "I'm done" at any point in this task. Once the participant elected to finish, the computer would thank him or her and recommend finding the research assistant for the final stage of the experiment.

Follow-up survey. The last stage of the experiment consisted of a survey for manipulation check (e.g., "How important did the color ranking task seem to you?"), questions about the design of the software (e.g., "How much effort did the computer expend in helping you on the desert survival task?"), and questions about the participant's general computer experience and knowledge (e.g., "How many years have you been using the computer?"). If the participant had not previously filled out the Mini Marker's survey (Saucier, 1994), then it was administered.

After the survey the research assistant debriefed the participant about the true nature of the experiment and the predetermined nature of the computer's behavior based on the randomly assigned condition. An explanation of the hypotheses was made, as well as a request to avoid telling any other students about the experiment in order to reduce bias in the data.

IVs and DVs. The independent variables of this experiment were the randomly assigned condition of the computer and the non-random rating of each participant's agreeableness and emotional stability states, as well as the gender of the participant. The dependent variable was the number of favors performed by the participant. Several components of the final survey were expected to be used as covariates, e.g., attribution styles (trust for the computer), and computer experience.

RESULTS

61 psychology students from the University of Utah took part in the study. Seven outliers were identified by the number of favors performed for the computer. These seven performed a very large number of favors, well beyond two standard deviations of the total sample, and were determined to be representative of another population than the sample of interest. Of the 54 participants included, the mean age was 24.5 years (M = 24.5, SD = 8.5). There were 25 women and 29 men.

The three statistics used throughout this analysis are the participants' agreeableness (M = .39, SD = .34), emotional stability (M = -.50, SD = .23) and number of favors performed for the computer (M = 10.93, SD = 11.48). There are four groups analyzed, which include the total set of participants (N = 54), the set of participants who worked with a helpful computer (N = 24), and the set of participants who worked with an unhelpful computer (N = 30).The asymmetry of these subgroups is due to the random assignment of the participants to each condition. The final group analyzed is the subset of participants who could be rated on the continuum of kin to reciprocal altruism (N = 26). The selection of this final group will be described below. It should also be noted that there was a significant correlation between agreeableness and gender ($r(52) = .33$, $p = .015$), so gender is included in the analyses.

Helpfulness of Computer

Of interest and concern was our inability to replicate the results found by Fogg in his original study. We first looked for a simple correlation between the helpfulness of the computer and the number of favors performed by the participants. There was no significant relationship between these variables. To explore the predictive power of helpfulness further, we performed a linear regression which controlled for gender, age, agreeableness, and ES. A significant prediction for number of favors, even holding these other variables constant, was not found.

Agreeableness

When taken as a whole, agreeableness could not be shown to correlate to the number of favors performed. When the subgroups were analyzed, however, agreeableness did correlate to the number of favors performed in the helpful computer condition ($r(27) = -.42$, $p = .041$), and did not correlate to the number of favors performed in the unhelpful condition. Using hierarchical regression, the predictive power of agreeableness in the helpful computer group was even stronger ($b = -20.63$, $t(19) = -2.21$, $p = .040$) showing that controlling for gender, age, and emotional stability, a one unit change in agreeableness predicted 21.6 fewer favors performed for the computer. There was no such predictive power for the unhelpful computer condition.

Emotional Stability

Taken as a whole, and also for each condition of helpfulness, our statistics were not able to show a correlation between ES and the number of favors performed for the computer, neither through a Pearson's r nor within a hierarchical regression.

Kin to Reciprocal Altruism

Ashton's 1998 research showed that the Saucier Mini-Markers could be parsed to create a scale of kin to reciprocal altruism. By selecting only the top half of the agreeableness scores, then using ES as a template for kin altruism (on the low end of the scale) and reciprocal altruism (on the high end of the scale), a subset of the sample (N = 26) was identified which could reasonably be used to represent a population of kin to reciprocal altruists. Because ES scores are independent from agreeableness scores, the descriptive statistics of kin to reciprocal altruism were re-centered with the mean at zero. This way, the range of scores went from -.56 on the low end to .56 on the high end. Kin altruists would be said to score on the low end of the continuum, and reciprocal altruists on the high end, with a mean score of -.03 and standard deviation of .25.

With this scale in mind, the correlation between kin to reciprocal altruism and number of favors performed was significant ($r(24) = -.46$, $p = .018$). This correlation is measured combining helpful/unhelpful conditions, meaning that the altruism style of the participant correlated to the number of favors performed, regardless of whether or not the computer was helpful. Indeed, breaking the groups apart we were not able to show a significant relationship between altruism style and number of favors in either case, while both trended downward like the meta-score.

An even stronger prediction for number of favors existed when altruism style was modeled within hierarchical

regression (b = -19.25, $t(21)$ = -2.65, p = .015), controlling for age, gender, and helpfulness of the computer. Altruism style also explained a significant proportion of variance in number of favors (R^2 = .26, $F(1, 21)$ = 7.00, p = .015). Our covariates involving computer familiarity and attribution styles did not show any effect throughout any of these analyses.

DISCUSSION

This study explored and expanded the CASA paradigm that states that people treat computers identical to people when those computers are imbued with specific social cues, even though the people they tested generally reported skepticism at the idea that they might treat the computer as anything other than a tool. If people do in fact act as if they implicitly believe a computer to be worthy of reciprocal benefit, then it was our goal to find out if people might reciprocate toward a computer in a manner similar to the way they reciprocate to another human and if personality predicts such reciprocations. The first of several surprises in our results was that we could not replicate the results of the original study by Fogg (1997). A helpful computer had no more of an effect on the number of favors it could elicit than an unhelpful computer could. As in any research, there is always the possibility of error this study, and it may be the case that the design simply failed to capture the phenomenon. We must also consider the possibility, however, that the effect has changed from the days 15 years ago when it was originally observed. There is support for this premise.

One explanation for our inability to replicate is the fact that computer technology has changed drastically in the past 15 years. It is less expensive, more ubiquitous, much more graphically appealing, and most of all, familiar to virtually all of the students who participated in our study. College-aged students today have grown up with computers in a way that those of 1997 did not. Whatever attributions these participants may make about intelligent software agents, it is almost certainly different than those attributions made by the previous generation.

A second reason for believing the original results have changed is that there were differences among the groups. Even though the helpfulness of the computer did not predict the number of favors performed, the personality measure, in certain cases, *did* predict the number of favors performed. If there were no attributions about the computer's worth as an object of altruism, then one would expect that no normative population of participants should show any difference in the number of favors performed. But this is not what happened.

When we analyzed the behavior of our sample while in the helpful computer condition, those who were low in agreeableness performed significantly more favors for the computer than those who were high in agreeableness. This was in direct contrast to our hypothesis. If people were treating computers by the same social conventions with which they treat other people, we would have expected to see a reversal of this effect. One possible explanation for this result is people who are high in agreeableness have only been shown to be more altruistic in experiments which involved an explicit interaction with another human being. Those same people may be unmotivated to reciprocate toward a computer than they are a human. On the other end of the spectrum, people who would shy away from reciprocating toward a human in an experimental setting are perhaps more comfortable when the object of benefit is a computer. We can only speculate as to why this distinction exists, but further research would be justified.

Finally, in our analysis of how people might treat a computer as either a friend or a stranger, the higher a participant ranked in kin altruism the more likely he or she was to perform a larger number of favors for the computer, regardless of whether or not the computer was helpful. This was also in contrast to our initial hypothesis, and raises the question of why people might imagine themselves to be in a communal (vs. exchange) relationship with an unfamiliar computer. One possible explanation may relate to the probability that the interface environment of a PC is highly familiar to the majority of computer users. It may be the case that a computer with a familiar operating system activates a set of expectations in the user which are independent of the setting and the machine. Given the ubiquity of modern computers, this set of expectations may be operated dozens of times per day for most people. When a powerful schema such as this is activated, it could possibly correlate with (or confound) measures which otherwise predict the judgments people make with respect to their intimate community.

This idea, while untested, gives new direction for future research into social HCI. There are several additional directions the research could go from this point. A human interactant could replace the computer. An older computer system paired with an older population of participants may ultimately replicate the results from the 1997 Fogg study. These research methods should be employed as we continue to explore whether or not the CASA paradigm has shifted.

References

Ashton, M. (1998).Kin Altruism, Reciprocal Altruism, and the Big Five Personality Factors.*Evolution and Human Behavior*, *19*(4), 243-255. doi:10.1016/S1090-5138(98)00009-9

Clark, M. S., & Mills, J. (1979).Interpersonal attraction in exchange and communal relationships.*Journal of Personality and Social Psychology*, *37*(1), 12-24. doi:10.1037//0022-3514.37.1.12

Eisenberger, R., Cotterell, N., & Marvel, J. (1987).Reciprocation ideology.*Journal of Personality and Social Psychology*, *53*(4), 743-750. doi:10.1037//0022-3514.53.4.743

Fogg, B. J. (1997). Charismatic computers: creating more likable and persuasive interactive technologies by leveraging principles from social psychology. Stanford University, Stanford, CA, USA.

John, O. P., Donahue, E. M., &Kentle, R. L. (1991).The Big Five Inventory--Versions 4a and 54.Berkeley, CA: University of California, Berkeley, Institute of Personality and Social Research.

John, O. P., Naumann, L. P., & Soto, C. J. (2008). Paradigm shift to the integrative Big Five trait taxonomy: History, measurement, and conceptual issues. In O. P. John, R. W. Robins, & L. A. Pervin (Eds.), Handbook of personality: Theory and research (pp. 114-158). New York, NY: Guilford Press.

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

Saucier, G. (1994). Mini-markers: A brief version of Goldberg's unipolar Big-Five markers. *Journal of personality assessment*. Retrieved from http://www.tandfonline.com/doi/abs/10.1207/s15327752jpa6303_8