

Is the Media Equation a Flash in the Pan? The Durability and Longevity of Social Responses to Computers.

Laura M. Pfeifer, Timothy Bickmore
College of Computer and Information Science
Northeastern University
360 Huntington Avenue, Boston, MA 02115
{laurap, bickmore}@ccs.neu.edu

ABSTRACT

Research on social responses to computers often assesses only first-impression reactions during a single experimental session, providing limited knowledge about the lasting effect of the results. In this work, we assess the lasting strength of social desirability bias effects on an interface designed to track exercise, manipulated to have high or low personalization (text vs. anthropomorphic conversational character). After 40 days of daily interactions by 25 participants, we found that self-reported exercise was more accurate when reported to the character vs. text. We also find that, for both conditions, participants' decision to initiate a session is greater when they have done more exercise. Moreover, we show that this effect significantly increases over time for participants in the character condition, and decreases for participants in the text condition. This study demonstrates that Media Equation effects can grow stronger or weaker over time, depending upon the presentation of the interface.

Author Keywords

Media Equation, self-report, longitudinal assessments

ACM Classification Keywords

H5.2. Information interfaces and presentation (e.g., HCI): User Interfaces—Evaluation/methodology, Graphical user interfaces, Interaction styles.

General Terms

Experimentation, Human Factors

INTRODUCTION

Since the introduction of the Media Equation hypothesis and the Computers as Social Actors paradigm, there have been several experiments showing that people respond socially to computers, in ways that are similar to how people respond socially to humans [12]. For example, people have been shown to assign gender and personality traits to computers and interact in ways that are biased by

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

CHI 2011, May 7–12, 2011, Vancouver, BC, Canada.

Copyright 2011 ACM 978-1-4503-0267-8/11/05...\$10.00.

those social assignments [9,12].

With years of research behind us, there is scientific support for the Media Equation hypothesis. However, research studies on the social responses to computers have typically only assessed a person's first-reaction response, and have not evaluated the potential for changes over time. A reasonable hypothesis is that when people interact with a social interface repeatedly, over long periods of time, their reactions to the interface might change. They may realize that the interface is not as capable or competent as a human is at being a social actor, or they may become conscious of their misattribution and as a result, their social responses to the computer may diminish with time. In this work, we explore the durability and longevity of social responses to computer interfaces, in a longitudinal context.

Social Desirability Bias

During social interactions, it is well-established that humans continually edit themselves to create a desirable presentation to their peers [5]. In survey or interview responses, a well-known form of self-editing is referred to as social desirability bias. This bias in self-reported data typically involves the over-reporting of socially desirable behaviors (voting, exercise, etc) and under- or non-reporting of socially undesirable behaviors (illicit drug use, etc) [4]. For example, self-reported information from online dating profiles is frequently inaccurate, with members describing themselves in a more socially positive light, e.g. men reporting themselves as taller and women reporting themselves as thinner [6].

When eliciting self-reports of potentially sensitive information, a well-known mechanism to reduce the effects of social desirability bias is to make the data collection medium as anonymous as possible. For example, computerized interviews have been shown to collect higher quality data than face-to-face interviews [2,10].

Nonetheless, even computerized interviews can elicit socially desirable responses. A classic study by Nass, et al., shows that people give significantly more polite responses when completing an evaluation about a computer, if the evaluation takes place on the same computer they recently interacted with, versus a different computer in the same room [8].

However, simple personalization manipulations of computer interfaces do not always elicit strong social cues. In an experiment examining the personalization of questionnaires and the corresponding effects of social desirability bias on self-reported data, Couper, et. al. failed to demonstrate socially biased reports on sensitive topics when elicited via a web survey that included a photo of the researcher, compared to the same survey without this personalization [13].

Cassell & Miller propose using embodied conversational agents – animated characters that simulate face-to-face interaction - as a medium for boosting perceptions of personalization, manipulating social bias effects, and for systematically studying self-report behaviors, since they encompass features of both a computer and a human [3]. They have the ability to build rapport, trust and accountability with the interviewee, as well as provide a sense of anonymity. Initial research has shown that people do give socially desirable responses to these high-personalization computer interfaces, at least during an initial session with the character [7].

In this paper, we extend previous work by exploring how social desirability bias effects, as witnessed in human-computer interactions, hold up over time. We describe a randomized longitudinal experiment, in which participants completed an assessment about their exercise behavior (walking), each day, for six weeks. The assessment was either conducted through dialogue with an anthropomorphic conversational character or via displayed text (Figure 1). We compare these self-report data to data collected via a wearable sensor (pedometer) worn by participants throughout the course of the study.

EXPERIMENTAL METHODS

The experiment was conducted in our Virtual Laboratory, a framework for running longitudinal experiments with a standing group of study participants [1]. For the current implementation, participants interacted with a computer character designed to emulate an exercise counselor and promote walking behavior. The daily conversation with the virtual exercise counselor took place on participants' home computers and lasted approximately 10 minutes. It included a discussion about the participant's walking behavior, and problem-solved any barriers to exercise. After the conversation with the virtual counselor was over, participants were given the opportunity to complete a single-item assessment, self-reporting the number of minutes they walked on the previous day. We chose this self-report measure as the target of this study because of its ability to be independently verified through the use of a wearable sensor: a pedometer.

Participants were randomized into one of two conditions, either completing the assessment via a continued conversation with their virtual counselor (CHARACTER, high personalization) or via displayed text (TEXT, low

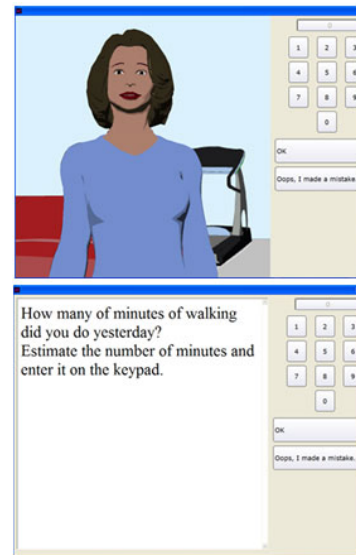


Figure 1. CHARACTER and TEXT condition.

personalization). All participants received the exercise counseling session with the character, and at the end of the interaction participants were asked - either spoken by the virtual counselor (CHARACTER) or via displayed text (TEXT) - “Do you have time for one more question?” If the participant responded *yes*, the participant was asked “How many minutes of walking did you do yesterday?” and the system displayed a keypad for the participant to enter their answer (Figure 1). This completed the participant's daily interaction with the system.

We independently measured walking behavior through the use of an Omron HJ-720ITC pedometer worn by each participant throughout the course of the experiment, with steps uploaded to the computer via USB cable. The pedometers displayed daily step counts but did not display any information regarding time spent walking.

Our hypotheses are that: (H1) the more physical activity the participants performed, the more likely they will be to report it to the system - and that this effect will be greater when reporting to the virtual human versus the displayed text; but that (H2) the effect will dissipate over time, as participants receive increased exposure to the computer system; and finally, (H3) data gathered via the character interface will be more accurate (verified by data from the pedometer) than data gathered via the text interface.

Participants

A total of 25 participants (23 female, ages 55-68) took part in the study. All participants were compensated \$1 per day that they interacted with the system. Thirteen participants were randomized into the CHARACTER condition and 12 were randomized into the TEXT condition. The experiment lasted for six weeks.

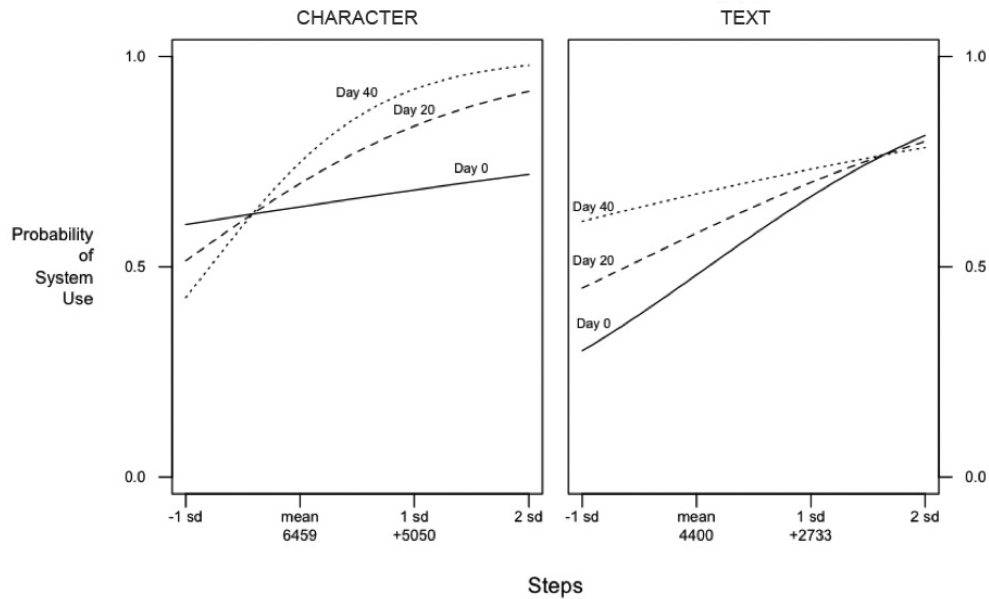


Figure 2. Effects of steps walked (pedometer-measured) on system use over time. Steps are mean-centered, by condition.

RESULTS

System usage and the self-reported walking behavior were analyzed as binary outcomes, by fitting a logistic mixed-effect regression model to the data. Logistic mixed-effect regression is a generalization of logistic regression, suitable for analyzing repeated binary measurements. All analyses were performed using R 2.9.1 with the lme4 package [11].

To examine H1 and H2, we first analyzed the likelihood of participants logging in and having a daily session with the system. Our model included fixed effects of study day, condition, and the pedometer step count from the previous day, along with corresponding interaction effects (Table 1). Results show a significant three-way interaction between the previous day's step count, study condition, and study day on system use. This interaction is visualized in Figure 2. Throughout the entirety of the study (days 0, 20 and 40 in Figure 2) we see that for both conditions, the more steps a participant walked, the more likely they were to use the system, providing partial support for H1. Initially, the effect is stronger in the TEXT condition than the CHARACTER condition.

Furthermore, we see significant differences between conditions on the lasting effect of this finding. By longitudinally examining the effect of the previous day's step count on daily system use, we find that in concordance with H2, this effect *decreased* over time for participants in the TEXT condition, but in contrast, it *increased* over time for participants in the CHARACTER condition. By the end of the study, the amount of walking done by a participant in the TEXT condition had little effect on their likelihood to use the system. In contrast, the amount of walking done by participants in the CHARACTER condition greatly determined the likelihood that they would interact with the system.

We also examined the likelihood of participants to self-report their minutes of walking once they initiated an interaction with the system. Our model included fixed effects of study day, condition, and the previous day's pedometer step count. Results show that participants in the TEXT condition were more likely to self-report their number of minutes walked, regardless of the number of steps they actually walked and regardless of the study day, compared to participants in the CHARACTER condition, $B = 1.817, p < 0.01$. We hypothesize this may be due to the TEXT interface being more efficient to use.

Table 1. Logistic Mixed-Effect Regression Estimate of Effects of Condition, Previous Day's Step Count, and Study Day on System Use. Condition 0 = CHARACTER, 1 = TEXT.

Random Effects	Std. Dev.		
Intercept	0.8457		
Fixed Effects	Estimate	Std. Error	<i>p</i>
Intercept	0.5858	0.3306	0.0764
Condition	-0.6632	0.4699	0.1581
Previous Day's Steps	0.1788	0.3055	0.5583
Study Day	0.0125	0.0103	0.2243
Cond*Steps	0.5917	0.4062	0.1453
Cond*Day	0.0074	0.0139	0.5940
Steps*Day	0.0300	0.0139	0.0306
Cond*Steps*Day	-0.0422	0.0172	0.0140

Finally, we examined the accuracy of the collected data. There was a significant positive correlation between pedometer step counts and self-reported minutes of walking, for both conditions, with a stronger correlation present for the CHARACTER condition, $r = 0.75$, $p < 0.001$; [95% CI, 0.69-0.81], vs. the TEXT condition: $r = 0.50$, $p < 0.001$; [95% CI, 0.40-0.59], providing support for H3.

CONCLUSION

Prior research indicates that people respond socially to computers, in similar ways to how they respond socially to other humans. In this work, we examine the durability of these effects in a longitudinal context. We show that social responses to computers change over time, and moreover, that this change varies depending upon the personalization of the interface. In a system designed for tracking exercise, we find that participants' actual walking behavior influences their daily decision to use the system. The more steps a participant walks, the more likely she is to use the system and conversely, the less she walks, the less likely she is to use the system. Furthermore, we find that with repeated interactions over time, this effect grows *smaller* for participants using a text-based interface and grows *larger* for participants using a character-based interface.

This study also examines the quantity and quality of data reported to the system. We find that participants are more likely to provide data to the text-based system, but that data provided to the character-based system were more accurate. It is possible that the character interface elicited higher-quality data due to increased feelings of trust and rapport, or because the time associated with reporting to the character - conversing with the character takes longer than reading text - caused participants to reflect more deeply about their answer, prior to reporting it to the system. Further research is needed to examine the exact reasons behind this finding.

One limitation of this study is that the experimental population is heavily biased towards women and older adults. Despite these limitations, this study provides support for the hypothesis that social desirability bias effects witnessed in human-computer interactions are not static, but can increase or decrease over time, and that these changes are influenced by the presentation of the computer interface.

ACKNOWLEDGMENTS

We thank Jenna Zaffini her assistance with this experiment and Daniel Schulman for his statistical insights. This work was supported by NSF CAREER IIS-0545932.

REFERENCES

1. Bickmore, T., Schulman, D., and Yin, L.: Maintaining Engagement in Long-term Interventions with Relational Agents. *International Journal of Applied Artificial Intelligence* 24, 6 (2010) 648-666.
2. Booth-Kewley, S., Larson, G. E., and Miyoshi, D. K.: Social desirability effects on computerized and paper-and-pencil questionnaires. *Computers in Human Behavior* 23, 1 (2007) 463-477.
3. Cassell, J. and Miller, P.: Is it Self-Administration if the Computer Gives you Encouraging Looks? In F. G. Conrad and M. F. Schober, Eds.: *Envisioning the Survey Interview of the Future*, John Wiley & Sons, (2007) 161-178.
4. Crowne, D. and Marlowe, D., *The approval motive; studies in evaluative dependence*, Wiley, (1964).
5. Goffman, E., *The presentation of self in everyday life*, Harmondsworth, (1978).
6. Hancock, J. T., Toma, C., and Ellison, N.: The truth about lying in online dating profiles. *Proceedings of the SIGCHI conference on Human factors in computing systems* (2007) 452.
7. Hoffmann, L., Krämer, N., Lam-chi, A., Kopp, S., Ruttkay, Z., Kipp, M., Nijholt, A., and Vilhjálmsson, H.: Media Equation Revisited: Do Users Show Polite Reactions towards an Embodied Agent? In *Intelligent Virtual Agents, Volume 5773*, Springer Berlin / Heidelberg, (2009) 159-165.
8. Nass, C., Moon, Y., and Carney, P.: Are People Polite to Computers? Responses to Computer-Based Interviewing Systems. *Journal of Applied Social Psychology* 29, 5 (1999) 1093-1109.
9. Nass, C., Steuer, J., and Tauber, E.: Computers are social actors. *CHI '94: Proceedings of the SIGCHI conference on Human factors in computing systems* (1994) 72-78.
10. Newman, J., Des, Turner, C., Gribble, J., Cooley, P., and Paone, D.: The Differential Effects of Face-to-Face and Computer Interview Modes. *Am J Public Health* 92, 2 (2002) 294-297.
11. R. Development Core Team, *R: A Language and Environment for Statistical Computing*, (2009).
12. Reeves, B. and Nass, C., *The media equation: how people treat computers, television, and new media like real people and places*, CSLI Publications, (1996).
13. Tourangeau, R., Couper, M., and Steiger, D.: Humanizing self-administered surveys: experiments on social presence in web and IVR surveys. *Computers in Human Behavior* 19, 1 (2003) 1-24.