

Ethical Issues in Using Relational Agents for Older Adults

Timothy Bickmore

College of Computer and Information Science
Northeastern University
Boston, Massachusetts
bickmore@ccs.neu.edu

Abstract

In this paper I outline some of the ethical arguments that have been made against anthropomorphic interfaces in general and those recently used in health behavior change interventions with older adults in particular, and I present some rebuttals to these arguments.

Introduction

Relational agents are computational artifacts designed to build and maintain long-term, social-emotional relationships with users (Bickmore 2003). In my work, I have focused on using embodied conversational agents for this purpose so that all of the verbal and nonverbal behavior that humans use face-to-face conversation can be employed for the purpose of conveying affective, attitudinal and relational information to users (Cassell, Sullivan et al. 2000). I have been particularly interested in the use of relational agents for health behavior change interventions in which both long-term interactions and strong working relationships between agents and users (referred to as “therapeutic alliances” in the medical literature) are required to maximize efficacy.

In two recent studies in which a relational agent played the role of an exercise advisor I was able to demonstrate that the use of relational behaviors by the agent led to higher working alliance scores, and that subjects who used the agent performed more physical activity than subjects in non-intervention control groups (although I have yet to be able to demonstrate a significant effect of relational behaviors on health behavior) (Bickmore, Caruso et al. to appear; Bickmore and Picard 2005).

The second of these studies involved 21 urban, mostly minority older adults aged 62 to 84, half of whom interacted with the agent daily for two months in their homes (see Figure 1). Of these subjects, seventeen (77%) were overweight or obese (based on BMI from reported height and weight), nineteen (86%) were scored as having low reading literacy, (Lobach, Hasselblad et al. 2003) eight (36%) reported never having used a computer before

and another six (27%) reported having used one “a few times”.

Acceptance and liking of the agent were rated highly by subjects in the intervention group, as was desire to continue using the system at the end of the intervention, with intervention subjects giving this an average rating of 6.2 (with 1=“not at all” and 7=“very much”).

“I enjoyed coming in and turning it on, and talking with her.”

The system used synchronized synthetic speech and animation for agent output, but limited user input to multiple choice selection of utterances using a touch screen. Participants felt that this simulated conversation worked reasonably well: when asked if they felt that they and the agent understood each other, participants rated this at 5.4 (with 1=“not at all” and 7=“very much”). However, several participants mentioned that they could not express themselves completely using this constrained form of interaction:

“When she ask me questions ... I can't ask her back the way I want”.

Thus, these initial results suggest that such systems may be well received and effective at changing health behaviors. However, it is important to ask whether the use of these systems is ethical; whether both the means and the ends are morally justifiable.

Some Ethical Issues

There are several ethical issues that are commonly raised regarding the use of anthropomorphic agents in general and the use of these agents for health behavior change in particular. These issues are particularly salient given that the older adult population is often considered to be in need of special protection.

One of the most common criticisms of this type of research is that users are deceived into thinking they are interacting with a person, and this deceit is unnecessary since the same positive outcomes could be achieved with a non-anthropomorphic interface. Let us address the two parts of this criticism separately. First, in this study, users clearly did not believe they were talking to a person when interacting with Laura. The experimenters never suggested this—Laura was introduced as a “cartoon

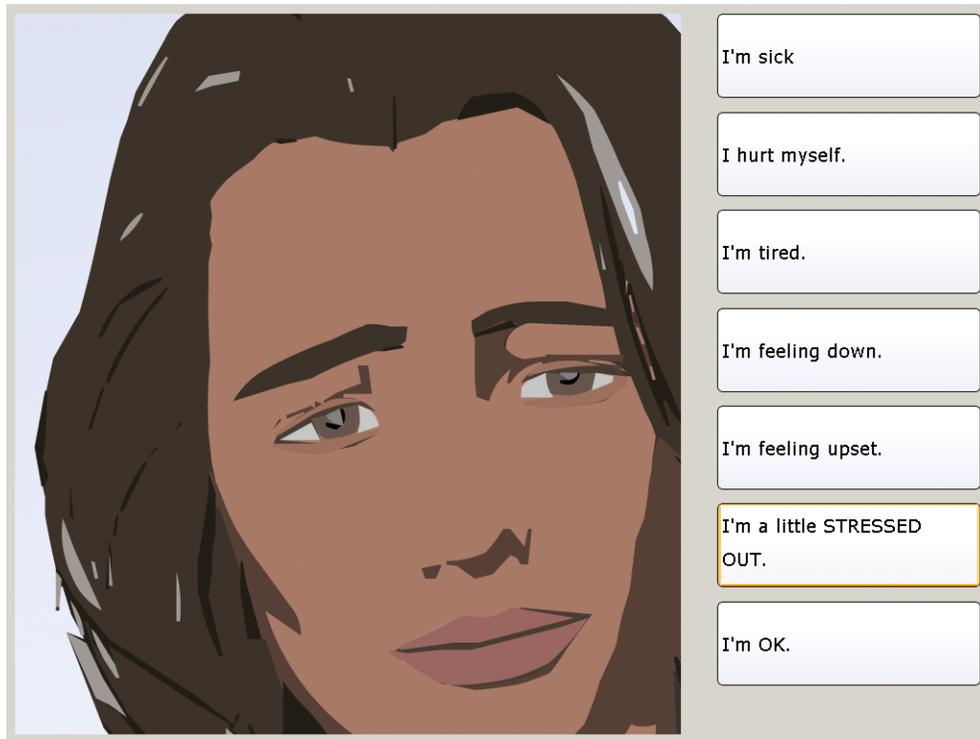


Figure 1. Relational Agent Interface was Accepted by Older Adults

character”—and Laura also periodically reminded subjects that she was “just a computer character with limited capabilities.” Subjects—with the exception of a few moments of forgetting—classified her somewhere between a machine and a person. Next, the contention that the same effects could be achieved with a conventional (non-relational) interface has a significant amount of evidence against it, at least for helping applications such as the one presented here, given the many studies that have shown the importance of working alliance, empathy and patient-centered communication for maximizing outcomes in long-term helping situations. These behaviors require some elements of an anthropomorphic interface to implement, including at least language and at most some form of body for nonverbal behavior. As with any interface comparison, however, this is ultimately an empirical question requiring a formal study to contrast the two specific interfaces—anthropomorphic and conventional—in question.

Other criticisms of agent-based interfaces have been made by Lanier and Shneiderman. Lanier is concerned that autonomous agents “dumb us down” by restricting our range of expressivity and action (Lanier 1995). There is some truth to this concern given the current interface modality that restricts users to multiple-choice selection of pre-written conversational utterances. However, one of my research goals is to find ways to allow the user greater expressivity while not raising expectations about the system’s functionality. In my opinion, presenting users with an unconstrained speech or free text input capability when the system can only respond in a small number of

ways in a given context is itself deceitful. Shneiderman’s criticisms are that anthropomorphic interfaces undermine predictability and consistency (leading to increased user anxiety), and that they have been a commercial failure (Shneiderman 1995; Shneiderman 1997). Regarding Shneiderman’s concerns, I feel that an embodied conversational interface can actually *enhance* users’ perceptions of predictability and consistency by providing a single interface modality across a wide range of functions, and which leverages their pre-existing knowledge and skills (i.e., of how to conduct face-to-face conversation) as much as possible. If anything, the users wanted *more* variability and novelty in the interface, not less. Perhaps the reason that anthropomorphic interfaces have not been a commercial success is that they have been repeatedly used in applications in which efficiency and accuracy are the most important requirements (e.g., as in banking machines). In one of Shneiderman’s critiques of anthropomorphic interfaces, he states that they are “cute the first time, silly the second time, and an annoying distraction the third time” (Shneiderman 1995). Interestingly, I have found exactly the opposite to be true in my studies for most users, even though contact with the agent did tend to taper off over time.

I would also argue that helping and psychotherapeutic applications such as the one presented here *require* the use of natural language in order to understand as much as possible about users’ thoughts, feelings and behavior and to deliver appropriately tailored interventions, and that as soon as an interface uses natural language in any form (including text) it is, in effect, an anthropomorphic

interface subject to all of the criticisms described above. One cannot present an utterance to a user (even brief text messages) without them making judgments about the personality, interpersonal attitude, relational stance and emotional state of the agent delivering the message (see Reeves & Nass for several examples (Reeves and Nass 1996)). Thus it is impossible to build a non-trivial system for this kind of application that is *not* anthropomorphic.

The ethics of deploying technology to change user behavior ('programming the user') is another area of concern. I argue that the issues here are no different than those faced by a human helper who is interested in changing a client's health behavior. The system can be said to be respecting the user's autonomy as long as: 1) the user is free to decide whether to use the system or not; 2) the user is free to decide whether to follow the system's recommendations or not; 3) the role of the system is primarily to provide information to enable the user to make informed choices; and 4) appropriate amounts of persuasion are used by the system only in cases in which the user is clearly making an unhealthy decision (O'Connell and Price 1983). Regarding this last point, while it is true that human helpers may be more understanding and adaptive than an agent (for now at least), the agent follows rules that are open to inspection and validation, and follows them in a consistent and unbiased manner, something that cannot always be said of human helpers.

Finally, I think that the results speak for themselves. Not only did the study involving older adults demonstrate significantly more walking behavior in the relational agent group, but I also found that all subjects enjoyed the overall experience and most enjoyed interacting with Laura on an ongoing basis. As the most vocal critic of Laura put it:

"It was the best thing that happened to me, to have something that pushed me out and get me walking."

Certainly from a utilitarian ethical framework at least, this kind of intervention is well justified.

Conclusion

Based on my experience with older adults, I believe that the caring, social, anthropomorphic interface enabled the exercise promotion system to be readily accepted, usable with minimal training, and effective in achieving a desired behavioral outcome with this user population. While the potential for misuse certainly exists (as with all tools), the potential for achieving beneficial outcomes has been demonstrated and represents a promising direction for future research.

Acknowledgements

Thanks to Jennifer Smith, Rosalind Picard, Ramesh Farzanfar, Jeffrey Mignault, and Robert Friedman for their many helpful comments on this article.

References

- Bickmore, T. 2003. Relational Agents: Effecting Change through Human-Computer Relationships. *Media Arts & Sciences*. Cambridge, MA, MIT.
- Bickmore, T., L. Caruso, et al. to appear. "It's just like you talk to a friend" - Relational Agents for Older Adults. *Interacting with Computers*.
- Bickmore, T. and R. Picard. 2005. Establishing and Maintaining Long-Term Human-Computer Relationships. *ACM Transactions on Computer Human Interaction* 12(2): 293-327.
- Cassell, J., J. Sullivan, et al., Eds. 2000. *Embodied Conversational Agents*. Cambridge, MA, The MIT Press.
- Lanier, J. 1995. Agents of Alienation. *Interactions of the ACM* 2(3): 66-72.
- Lobach, D., V. Hasselblad, et al. 2003. *Evaluation of a Tool to Categorize Patients by Reading Literacy and Computer Skill to Facilitate the Computer-Administered Patient Interview*. AMIA, Washington, DC.
- O'Connell, J. and J. Price. 1983. Ethical Theories for Promoting Health Through Behavioral Change. *J of School Health* 53(8): 476-479.
- Reeves, B. and C. Nass. 1996. *The Media Equation*. Cambridge, Cambridge University Press.
- Shneiderman, B. 1995. Looking for the bright side of user interface agents. *interactions* 2(1): 13-15.
- Shneiderman, B. 1997. *Direct manipulation for comprehensible, predictable and controllable user interfaces*. International Conference on Intelligent User Interfaces, Orlando, FL.