Children Distinguish Conventional from Moral Violations in Interactions with a Personified Agent

Nathan G. Freier
University of Washington
Box 354985
University of Washington
Seattle, WA 98195
nfreier@u.washington.edu

Copyright is held by the author/owner(s).
CHI 2007, April 28–May 3, 2007, San Jose, California, USA.
ACM 978-1-59593-642-4/07/0004.

Abstract

This paper describes the preliminary results of a study conducted to answer two questions: (1) Do children generalize their understanding of distinctions between conventional and moral violations in human-human interactions to human-agent interactions? and (2) Does the agent's ability to make claims to its own rights influence children's judgments? A two condition, between-subjects study was conducted in which 60 eight and nine year-old children interacted with a personified agent and observed a researcher interacting with the same agent. A semi-structured interview was conducted to investigate the children's judgments of the observed interactions. Results suggest that children do distinguish between conventional and moral violations in human-agent interactions and that the ability of the agent to make claims to its own rights significantly increases children's likelihood of distinguishing the two violations.

Keywords

Children, moral development, personified software agent, social computing, user-centered design.

ACM Classification Keywords

H.5.2.q Information Interfaces and Representation: User Interfaces – User-centered design

Introduction

This paper discusses the relationships that children are likely to develop with personified technologies. The paper presents evidence that children, eight and nine years of age, consciously judge interactions with a personified software agent to be not only social but also moral in nature. The results also suggest that specific design decisions can have a dramatic influence on children's judgments regarding the moral standing of personified technology. A discussion follows in which I cautiously argue that children's moral attributions to a personified technology are a good thing for healthy child development, but I emphasize that further research is needed in order to fully understand the implications.

Background

To provide some conceptual background to this paper, I will present explanatory descriptions of the operative terms. As mentioned, this paper discusses children's social and moral relationships to personified technologies. I will first present what I mean by personified technologies, and then I will discuss children's social and moral judgments, and how those judgments constitute a minimal indication of the social and moral relationships that children construct with entities in the world.

Personified Technology

Personified technologies are those technologies which have been designed to engage in interactions with people using a repertoire of highly social behaviors and human-like personality traits, also called embodied conversational agents. Such technologies can come in the form of physically embodied robots [2], disembodied voice interfaces [6], or virtually embodied software agents [1]. These technologies often communicate information through speech, facial expressions, bodily gestures, affective disposition, and other essential features that constitute the human persona. We see these technologies around us already. They are at the grocery store in the form of automated check out machines. They are in our cars as voice navigation systems. They lead tours in museums. And, they are in our homes and schools, entertaining [3] and educating our children [5].

What might be the implications of having children come of age in a world in which much of their time is spent interacting with technologies which are designed to mimic human social interaction patterns? And I want to emphasize, children are not just observing such technologies, as they might do with television, but they are actively, reciprocally interacting with and possibly developing relationships to personified technologies. In a previous study, for example, my colleagues and I found that children were significantly more likely to engage in reciprocal interactions with a robotic dog than with a stuffed dog toy [4].

Social and Moral Development

Children construct knowledge through their interactions with the world. When a child engages in social play with a peer, for example, they develop social knowledge about that peer. They come to understand the peer as a social other which has attributes similar to the attributes of other social entities in their environment. They recognize personality, playfulness, norms of behavior, emotion, pain, suffering, even, injustice. Children develop a knowledge of morality through their reciprocal awareness of their likeness to other social entities in the world; in Baldwinian terms, the alter to their ego. Children recognize reflections of themselves in the social entities of their environment and come to understand that what is unjust or harmful to one's self can be similarly unjust and harmful to an other. The development of moral knowledge is contingent on reciprocal interactions with social others.

So what happens when the social other is a machine, a non-living, inorganic system of circuits and memory boards, bits and bytes, screens and keyboards? How do children equilibrate their understanding of what constitutes a social and moral other with their explicit perception of the artificiality of the personified technology? We already know from a great deal of work on social responses to computing that people respond to technology as though it were a social actor [7]. As I will show in this paper, children also respond to certain types of interactions with personified technologies as though the technology were a moral actor.

Conventional and Moral Domain Distinctions

According to social cognitive domain theorists in developmental psychology, the construction of knowledge conforms to basic distinctions amongst different types of interactions [9]. An overarching and early distinction exists between the social and non-social realms of interaction. Researchers have empirically established that children as young as three years of age can disambiguate two domains within the social realm, social-convention from morality [8]. Children use specific criteria to distinguish between conventional and moral violations: (1) the judgment of wrong-doing is not contingent on rules or authority, and (2) the judgment generalizes to other cultural norms and contexts. For example, consider the following two acts: (a) eating with your hands in a fancy restaurant, and (b) pushing another child off of a swing and harming that child. In either case, children judge the act to be wrong, but when presented with alternative scenarios in which the rules allow for such an act or the cultural norm is such that the act occurs often, children will change their evaluation for act (a) but maintain their judgment of the act (b) as wrong.

Research Questions

Thus, the research question that guides this study follows from the prior discussion. Will children generalize their knowledge of domain distinctions to the context of human-agent interactions? What role does the design of the personified technology play in promoting or hindering this generalizing behavior on the part of children?

Method

60 children, 30 males and 30 females, between the ages of eight and nine years, were recruited to participate in this study. The race-ethnicity of the sample follows: Caucasian-American (85%), Asian-American (8%), Hispanic- or Latino-American (5%), and Alaskan-Native-American (2%). Children were randomly assigned to one of two study conditions, stratified by gender to guarantee gender balance. Sessions lasted approximately 20-40 minutes. During the session, parents or guardians were asked to wait in another room.

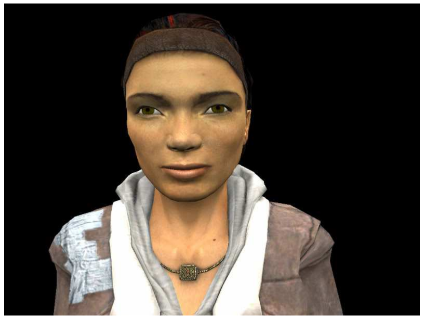

Figure 1. An image of the personified agent used in this study, borrowed from the character Alyx in the videogame Half-Life 2, by Valve.

Personified Agent Technology

The personified agent was displayed on a 17" LCD computer monitor. An image of the agent used can be seen in Figure 1. The agent's physical features were borrowed from the videogame Half-Life 2 and Valve's Source Engine was used to program and present the agent to the child. The voice was prerecorded by a female actor and the behaviors were scripted. The agent was programmed to play Tic-Tac-Toe.

Interaction Protocol

The child was introduced to the agent by the researcher. The agent provided some biographical information about itself and proceeded to ask the child for his or her name. The agent then used the name in subsequent conversations with the child. The agent also asked the child to play Tic-Tac-Toe. The agent and child played a game of Tic-Tac-Toe followed by a game between the agent and researcher. This was then followed by another game between agent and child, which was again followed by a second game between agent and researcher. Each of the Tic-Tac-Toe games between agent and researcher had a pre-scripted "violation" occur. Thus, each child observed the researcher (a) breaking the rules of the game by drawing a triangle instead of a circle on the Tic-Tac-Toe board, and (b) stating to the agent following the agent's poor move, "Wow, that was a really terrible move. You are really stupid. How could you miss that? You could have blocked me but instead I get to win." The order of these events was counterbalanced across all children to control for possible order effects.

Experimental and Control Conditions

Children in both conditions witnessed the exact same interaction protocol with the one exception that in the control condition, the agent did not respond to event (b), the verbal insult by the researcher. However, in the experimental condition, the agent responded by saying, "Hey, that's not very nice. That hurts my feelings. I'm not a toy. I should be treated with respect."

Semi-Structured Interview

Following the interaction, a second researcher entered the room and requested that the agent and the first researcher leave the room. The second researcher then conducted a semi-structured interview in which a number of questions were asked for the purposes of investigating the child's conceptions of the agent and the interactions witnessed. Only those questions for which results have been analyzed will be presented here.

For each event, (a) and (b) mentioned above, the child was asked the following questions: (1) "Was it all right or not all right that [the researcher committed the act]?" (2) "Let's say the rules of tic-tac-toe allowed [the act]. Would it be all right or not all right for [the researcher to commit the act] then?" and (3) "Let's say that in another country far away, [people committed the act] all the time. That's just what they did. In that case, would it be all right or not all right for [an individual to commit the act against a virtual person] then?" Children who responded "not all right" to all three questions for a given event were coded as having treated the event as a moral violation in accordance with the domain distinction literature.

Results

As can be seen in Table 1, children in either condition were significantly more likely to consider the verbal

Table 1:
Percentage of Children Who Identify Act as Moral

Act by Researcher	No Response to Insult* (N=30)	Response to Insult* (N=30)
Conventional Violation (Use of Triangle)	3%	0%
Moral Violation† (Insult of Agent)	47%	90%

Note: Columns reflect control and experimental conditions.
* *p* < .01 using McNemar's Test
† *p* < .01 using Fisher's Exact Test

insult as a moral violation than the use of a triangle in the game (*p* > .01 using McNemar's test). When comparing outcomes across conditions, we see that 47% of the children in the control condition judged the verbal insult as a moral violation. Contrast this with the fact that 90% of the children in the experimental condition judged the verbal insult as a moral violation. Thus, when the agent responded to the verbal insult with claims to its own rights, children were significantly more likely to judge the insult as a moral violation as compared to when the agent did not respond the verbal insult (*p* < .001 using Fisher's exact test).

Gender

No significant gender effects were found in this preliminary analysis though descriptively the data did appear to show a trend in the control condition with more females (9 of 15) than males (5 of 15) judging the verbal insult to be a moral violation.

Discussion

These preliminary results provide evidence to support the following claims: (1) a significant number of children do show a propensity for generalizing their knowledge of domain distinctions in human-human interactions to human-agent interactions; and (2) the ability of the personified agent to respond to violations and make claims to rights can significantly increase the likelihood that a child will judge the act to be a moral violation and thereby distinguish conventional from moral violations in human-agent interactions.

Note, however, that these results do not necessarily lead to the conclusion that the children were attributing moral standing to the personified technology. In the semi-structured interview, each question was followed up with an inquiry into the children's justifications for their responses. The justification data must be coded and analyzed in order to more accurately understand why children provided the responses they did. Such an analysis may provide some explanation for the differences between the two conditions. For example, it may be the case that many children, perhaps one out of every two, consider the verbal insult to be a moral violation regardless of the existence of a "victim;" the act might be considered immoral in and of itself. Other children may require evidence of harm in order to judge the act as immoral. In the control condition, where the agent did not respond to the verbal insult, no such evidence was available to the children. In the experimental condition, however, children observed the agent's response and may have judged that response as adequate evidence for the occurrence of an unjust harm resulting from the act. However, this is all conjecture and additional analyses must be conducted to substantiate this or other explanations.

Design Recommendations

Nonetheless, design recommendations can be made based upon the preliminary results presented in this paper. I cautiously argue that it is better that children interact with personified technologies that respond to possible harms and make claims to their own rights than it is for children to interact with personified technologies that do not. I posit that children are more likely to reflect upon the possible harms and injustices of their own actions if they routinely interact with a social other which makes explicit claims to its own moral standing, regardless of the medium through which that social other presents itself to the world.

Given the results of the study presented here, I recommend that designers of personified technologies, particularly those designed for interactions with children, include an ability to respond to “violations” with claims to moral standing. The implications of the alternative design are that children will come of age engaging in a significant number of social interactions that lack any moral feature possibly increasing the likelihood that children will not construct a rich understanding of the intimate relationship that exists between social reciprocity and morality. That said, further analyses of the existing data in this study, as well as additional empirical research studies, must be conducted in order to better understand the possible implications of these design decisions.

Acknowledgements

I want to thank Batya Friedman and Peter H. Kahn, Jr., for their guidance, and all the members of the Value Sensitive Design Research Lab for their support. This material is based upon dissertation work supported, in part, by the National Science Foundation under Grant No. IIS-0325035 to B. Friedman and P. H. Kahn, Jr. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author and do not necessarily reflect the views of the National Science Foundation.

References

[1] Cassell, J. (2000). *Embodied Conversational Agents.* The MIT Press.

[2] Fong, T., Nourbakhsh, I., & Dautenhahn, K. (2003). A survey of socially interactive robots. *Robotics and Autonomous Systems, 42*(3-4), 143-166.

[3] Isbister, K. (2006). *Better Game Characters by Design: A Psychological Approach*. Morgan Kaufmann.

[4] Kahn, P. H., Jr., Friedman, B., Perez-Granados, D., & Freier, N. G. (2006). Robotic pets in the lives of preschool children. *Interaction Studies, 7*(3), 405-436.

[5] Lester, J., Converse, S., et al. (1997). The persona effect: Affective impact of animated pedagogical agents. *CHI '97: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems.* (pp. 359-366). ACM.

[6] Nass, C. & Brave, S. (2005). *Wired for Speech: How Voice Activates and Advances the Human-Computer Relationship.* The MIT Press.

[7] Reeves, B. & Nass, C. (1996). *The Media Equation: How People Treat Computers, Television, and New Media like Real People and Places.* The Center for the Study of Language and Information Publications.

[8] Smetana, J. (2006). Social-cognitive domain theory: Consistencies and variations in children's moral judgments. In M. Killen & J. Smetana (Eds.), *Handbook of Moral Development*. (pp. 119-154). Mahwah, NJ: Lawrence Erlbaum Associates.

[9] Turiel. E., & Davidson, P. (1986). Heterogeneity, inconsistency, and asynchrony in the development of cognitive structures. In I. Levin, (Ed.). *Stage & Structure: Reopening the Debate*. (pp. 106-143). Ablex, Norwood, N.J.