In: Proceedings of the Workshop on Shaping Human-Robot Interaction:
"Understanding the Social Aspects of Intelligent Robotic Products."
In Cooperation with the ACM CHI'04 Conference, Vienna, Austria.

From Fiction to Science – A cultural reflection of social robots

Christoph Bartneck
Eindhoven University of Technology
Den Dolech 2, 5600 MB Eindhoven
The Netherlands
Christoph@bartneck.de

ABSTRACT

This paper reflects on the culture of human-robot interaction. A review of common concepts in movies and literature is presented and their relation to scientific work is discussed. Two new research directions on the synthesis of behavior models and the perception of social robots are presented.

Author Keywords
Social Robots, Society, Interaction

ACM Classification Keywords
Interaction design, robots, social products

INTRODUCTION

The role that robots might play in our society and what their abilities will be, has been an important topic in science fiction literature. Issac Asimov (1991) defined the Three Laws of Robotics which set a framework of human-robot interaction:

1. A robot may not injure a human being or, through inaction, allow a human being to come to harm.
2. A robot must obey orders given it by human beings except where such orders would conflict with the First Law.
3. A robot must protect its own existence as long as such protection does not conflict with the First or Second Law.

The arrivals of the first consumer robots, such as AIBO, confront us with the need to take these frameworks and ideas out of fiction and into reality. To start with, I would like to review common concepts in the science fiction domain.

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.
Conference'04, Month 1Ð2, 2004, City, State, Country.
Copyright 2004 ACM 1-58113-000-0/00/0004É$5.00.

ROBOTS WILL TAKE OVER THE WORLD

The movie ÒAnimatrixÓ (see Figure 1) describes the second renaissance as a period in which humanity created millions of robots to server their needs, the process of the robotÕs emancipation, the war of humanity against robots with the final stage of humanitiesÕ enslavement (Wachowski & Wachowski, 2003).

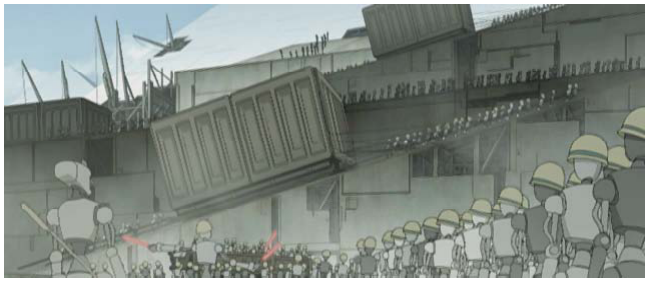

Figure 1: Scene from Animatrix

This is a typical scenario that can also be found in ÒTerminatorÓ (see Figure 2) and other Hollywood movies.

Figure 2: Terminator

Interestingly, this vision of the future is not so popular in Japan. A possible reason for this could be that the in Shinto Buddhism god is in everything, including humans, animals, plants and machines and rocks. The Christian world makes a strict division between creatures that have a soul and

objects that do not. According to Shinto, robots are not that different from humans. In the popular Japanese Manga movies good fights evil just like in the western world, but the role of the good and the evil is not mapped directly to humans as being the good against robots being the evil. In these movies the good and the evil are distributed. You might have a good robot that fights an evil human villain or a good robot fighting bad robots.

ROBOTS WANT TO BE LIKE HUMANS

In the TV series "Star Trek – Next Generation" the android Data (see Figure 3) is constantly trying to become more human (Paramount Pictures, 2002).

Figure 3: Data

At some point he even acquires an emotion chip that enable him to experience feelings. I cannot see a good reason for Data's behavior other than that the writers of the series wanted to flatter humanity. It is perfectly acceptable that he would want to be able to communicate effectively with the crew of the Enterprise and thus it makes sense that he studies their behavior. But why would Data want to become human? Why would he want to be something that he cannot be?

Figure 4: AI

The same assumption is present in Steven Spielberg's "AI" (see Figure 4) in which the main robot character wants to become "a real boy" (Watson & Aldiss, 2001) and in "Bicentennial Man" (Asimov, 1999).

Data's brother Lore does not have this need, but again the writers slip back to the "robots will take over the world" scenario. In the double episode "Descent" Lore teams up with the Borg to take over the universe's leadership from the inferior biological life forms. It appears difficult for humans to accept that other intelligent beings would not want to be like them. In particular if humans originally created these beings. The situation appears similar to the one that parents face when their children are very different from them and choose a different life style than themselves. - Those ungrateful (robot) brats do whatever they want! - Maybe we ourselves have to mature and let robots be what they really are or want to be. This will of course still have to be within the boundaries of the laws. Robots will have to respect them as any other member of the society.

In the animated TV series "Futurama" the robot Bender (see Figure 5) demonstrates what robot emancipation can be like (Groening, 1999).

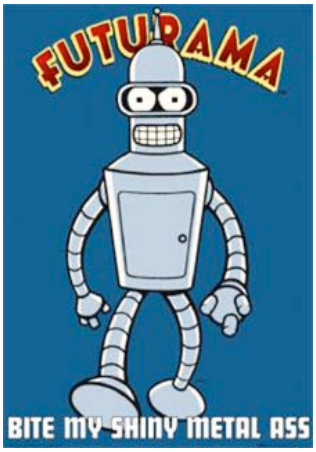

Figure 5: Bender

Bender and all the other robots live along humans, but are happy with what they are. Confronted with the choice to be a human Bender is most likely to answer: "bite my shiny metal ass!", which also illustrates Bender's general attitude.

PEOPLE WANT ROBOTS TO BE LIKE HUMANS

In the movie Star Wars: Episode V - The Empire Strikes Back (1980) by George Lucas the group of Jedi Nights and robots visit the City in the Sky. During their visit one of the main characters, the robot R2D2 (see Figure 6), separates from the group and enters a secret room.

To his own dismay, he entered a robot torture chamber. A similar model to R2D2 is turned up side down and glowing irons are pressed against his feed. The robot unsuccessfully wiggles to avoid the irons and upon contact beeps out loud. R2D2 is shocked and afraid and expresses his distress with a series of beeps. This whole scene makes no sense whatsoever. Robots cannot feel pain, do not have emotions

and torture is a very ineffective way to extract information from them. Still, the viewers feel sorry for the tortured robot and worries about R2D2.

Figure 6: R2D2

This movie scene demonstrates how easily people attribute human emotions into machines. One can conclude that robots only need to mimic human behavior as closely as possible to be perceived as a social being. Given this assumption the creators of robots have to formalize existing psychological models of human behavior, such as the OCC model (Ortony, Clore, & Collins, 1988) for an emotion system, and implement them into the robot. Since these models were usually only created to explain human behavior and not to synthesis it, it takes a considerable effort to convert them into a working software model (Bartneck, 2002).

RESEARCH DIRECTIONS

But this is exactly one of the most interesting areas of robot research: synthesizing human behavior to validate psychological models of human behavior. Synthesis is an essential activity in scientific conduct but psychologists were limited to the analyses of human behavior since they were not able to create an artificial being that they could use to synthesize human behavior. With the maturity of robotic technology, including a significant increase of computing power, it became possible to create intelligent social beings. We now can create robots that act autonomously in the real world and that interact with humans (Fong, Nourbakhsh, & Dautenhahn, 2003). By cycles of synthesis and analysis we will be able to create robots that act naturally with humans and at the same time gain a better understanding of humans themselves. The arrival of studies into the ethical (Dennet, 1997) and legal (Lehman-Wilzig, 1981) aspects of human-robot interaction shows that the integration of robots in our society in immanent.

All of this is still under the assumption that humans want robots to act human like. A very interesting area for robot research is the questions when do people not treat robots like humans. When does the perception of a social robot break down and the robot is treated like a machine. To find a clear answer to this question it appears necessary to take a closer look into more extreme situations, possibly negative ones. Here is a list of possibly research questions:

- Do humans torture robots differently than other humans?
- How do humans treat robots that lie and cheat?
- Do humans hold robots responsible for their failures?

By understanding human attributions in extreme conditions, we might discover effects that we could observe to a lesser degree in common human-robot scenarios. They might help us better understand the human-robot interaction.

CONCLUSIONS

It appears necessary to let go some concepts about human robot interaction that have been promoted by movies and literature. They utilize people's fear of the unknown to build engaging stories. These concepts are therefore so strong in the minds of the people that they might have influence on the results of our empirical studies, since the participants in our experiments are exposed to them too.

Two main research areas appear to be of interest. First, the synthesis of human behavior models based on existing models and their further development through iterative cycles. Second, the study of the perception of robots as social actors. When do people perceive them as social beings and when like machines and what is the influence of this perception on the interaction between them?

REFERENCES

1. Asimov, I. (1991). I, robot. New York: Bantam Books.
2. Asimov, I. (Writer) (1999). Bicentennial Man.
3. Bartneck, C. (2002). *Integrating the OCC Model of Emotions in Embodied Characters*. Proceedings of the Workshop on Workshop on Virtual Conversational Characters: Applications, Methods, and Research Challenges, Melbourne.
4. Dennet, D., C. (1997). When HAL Kills, Who's to Blame – Computer Ethics. In D. G. Storck (Ed.), *Hal's Legacy* (pp. 351-365). Cambridge: MIT Press.
5. Fong, T., Nourbakhsh, I., & Dautenhahn, K. (2003). *A survery of socially interactive robots*. Robotics and Autonomous Systems, 42, 143-166.
6. Groening, M. (Writer) (1999). Futurama.
7. Lehman-Wilzig, S. N. (1981). *Frankenstein Unbound: Toward a Legal Definition of Artificial Intelligence*. FUTURES: The Journal of Forecasting and Planning, 13(6), 442-457.

8. Ortony, A., Clore, G., & Collins, A. (1988). The *Cognitive Structure of Emotions*. Cambridge: Cambridge University Press.
9. Paramount Pictures. (2002). StarTrek. from http://www.startrek.com/
10. Wachowski, A., & Wachowski, L. (Writer) (2003). Animatrix.
11. Watson, I., & Aldiss, B. (Writer) (2001). AI.