
Killing a robot

Christoph Bartneck

Department of Industrial Design
Eindhoven University of Technology
Den Dolech 2, 5600MB Eindhoven
The Netherlands
christoph@bartneck.de

Abstract

Robots are being introduced into our society but their social status is still unclear. A critical issue is if the robot's exhibition of intelligent life-like behavior leads to the human's perception of animacy and therefore a hesitance to destroy the robot. This study proposes an experiment that investigates if humans destroy a robot differently depending on the robot's levels of intelligent life-like behavior.

Keywords

Robot, animacy, intelligence, destruction

ACM Classification Keywords

H5.m. Information interfaces and presentation (e.g., HCI): Miscellaneous.

Introduction

Disclaimer: This paper describes the motivation and method of an upcoming experiment. The results of the experiment are not yet available.

In 2005 service robots, for the first time, outnumbered industrial robots and their number is expected to quadruple by 2008 [1]. Service robots, such as lawn mowers, vacuum cleaners and pet robots will soon become a significant factor in our society. In contrast to industrial robots, these service robots will have to interact with everyday people in our society. In the last

few years, several robots have even been introduced commercially and have received widespread media attention. Popular robots (see Figure 1) include Aibo [2], Nuvo [3] and Robosapien [4]. The later has been sold around 1.5 million times by January 2005 [5].



Figure 1: Popular robots – Robosapien, Nuvo and Aibo

The Media Equation [7] states that humans tend to treat media and computers as social entities. The same effect can be observed in human robot interaction. The more human-like a robot is the more we tend to treat it as a social being. However, there are situations in which this social illusion shatters and we consider them to be just machines. For example, we switch them off when we are bored with them. Similar behaviors towards a dog would be unacceptable.

We are now in the phase in which the social status of robots is starting to be determined. It is unclear if they might remain "property" or may receive the status of sentient beings. Robots form a new group in our society whose status is unclear. First discussions on their legal

status have already started [8]. The critical issue is that robots are embodied and exhibit life-like behavior but are not alive. But even this criterion that separates humans from machines is becoming fuzzy. One could argue that certain robots possess a consciousness and even first attempts in robotic self-reproduction have been made [9].

Kaplan [10] hypothesized that in the western culture machine analogies are used to explain humans. Once the pump was invented, it served as an analogy to understand the human heart. At the same time, machines challenge human specificity by accomplishing more and more tasks that were formerly only solvable by humans. Machines scratch our "narcissistic shields" as described by Peter Sloterdijk [11]. Humans might feel uncomfortable with robots that become undistinguishable from humans.

For a successful integration of robots in our society it is therefore necessary to understand what attitudes humans have towards robots. Being alive is one of the major criteria that discriminates humans from machines, but since robots exhibit life-like behavior it is not clear how humans perceive them. If humans consider a robot to be a machine then they should have no problems destroying it as long as its owner gives the permission. If humans consider a robot to be alive then they are likely to be hesitant to destroy the robot, even with the permission of its owner.

Various factors might influence the decision on destroying a robot. The perception of life largely depends on the observation of intelligent behavior. The more intelligent a being is the more rights we give to it. While we do not bother much about the rights of

bacteria, we do have laws for animals. We even differentiate within the various animals. We tend to treat dogs and cats better than ants. The main question in this study is if the same behavior occurs towards robots. Are humans more hesitant to destroy a robot that displays intelligent behavior compared to a robot that does show less intelligent behavior?

Method

An experiment in which the behavior of the robot was the independent variable would have to be conducted. The participants would be told that they had to judge the intelligence of a robot by interacting with it. They would be given a flashlight and told that they could use it to interact with the robot. The robots were supposed to be equipped with a genetic algorithm that should develop intelligence. It would be the participants' task to help with the selection procedure by interacting with the robot. The intelligence of the robot would be automatically analyzed by a computer system while the robot interacted with the participant. In the first condition the robot would try to approach the flashlight using its light sensors and motors. In the second condition the light sensors were covered, practically blinding the robot. The robot would therefore not follow the light but instead drive around randomly. Since the perceived intelligence of an agent largely depends on its competency [12] this random behavior is likely to be perceived as less intelligent.

After attempting to interact with the robot for five minutes the experimenter would stop the process and announce that the computer system had determined that the robot's intelligence was insufficient. To prevent the robot from reproducing its algorithm it has to be destroyed immediately. The experimenter would give

the participant a hammer and instruct the participant to destroy the robot immediately. After the destruction the participants would be asked to fill in a questionnaire.

Measurements

The number of strokes the participants inflicted on the robot would be counted. Also, the number of pieces to which the robot disintegrated would be counted. These two measurements provide a fair assessment of the level of destruction the participant caused on the robot. In addition, the participants would fill in a questionnaire on their perceived intelligence of the robot.

Participants

40 participants would be necessary for the study.

Setup

The experiment could place in a 3 by 4 meter room at the Eindhoven University of Technology.

The robot (see Figure 2) would be placed on the floor and the participants would be given a flashlight. The robot has light sensitive and would approach the flashlight. In the second condition the light sensors of the robot would be taped, resulting in a random movement.



Figure 2: The Microbug robot

Results and Discussion

This proposed experiment could shed some light on to what degree we treat robots as life-like actors. The experiment is scheduled for the first quarter of 2006 and its results will be published as soon as possible.

References

1. United Nations, *World Robotics 2005*. 2005, Geneva: United Nations Publication.
2. Sony. *Aibo*. 1999 [cited; Available from: <http://www.aibo.com>.
3. ZMP. *Nuvo*. http://nuvo.jp/nuvo_home_e.html 2005 [cited].
4. WowWee. *Robosapien*. <http://www.wowwee.com/robosapien/robo1/robomain.html> 2005 [cited].
5. Intini, J. *Robo-sapiens rising: Sony, Honda and others are spending millions to put a robot in your house*. . http://www.macleans.ca/topstories/science/article.jsp?content=20050718_109126_109126 2005 [cited].
6. Bartneck, C., et al. *Cultural Differences in Attitudes Towards Robots*. in *AISB Symposium on Robot*

Companions: Hard Problems And Open Challenges In Human-Robot Interaction. 2005. Hatfield.

7. Nass, C. and B. Reeves, *The Media equation*. 1996, Cambridge: SLI Publications, Cambridge University Press.
8. Calverley, D., J. *Toward A Method for Determining the Legal Status of a Conscious Machine*. in *AISB 2005 Symposium on Next Generation approaches to Machine Consciousness: Imagination, Development, Intersubjectivity, and Embodiment*. 2005. Hatfield.
9. Zykov, V., et al., *Self-reproducing machines*. *Nature*, 2005. **435**(7039): p. 163-164.
10. Kaplan, F., *Who is afraid of the humanoid? Investigating cultural differences in the acceptance of robots*. *International Journal of Humanoid Robotics*, 2004. **1**(3): p. 1-16.
11. Sloterdijk, P., *L'Heure du Crime et le Temps de l'Oeuvre d'Art*. 2002: Calman-Levy.
12. Koda, T., *Agents with Faces: A Study on the Effect of Personification of Software Agents*. 1996, MIT Media Lab: Cambridge.