“Daisy, Daisy, Give Me Your Answer Do!”
Switching Off a Robot

Christoph Bartneck, Michel van der Hoek
Department of Industrial Design
Eindhoven University of Technology
Den Dolech 2, 5600MB Eindhoven, The Netherlands
+31 40 247 5175
christoph@bartneck.de
m.j.v.d.hoek@student.tue.nl

Omar Mubin, Abdullah Al Mahmud
User-System Interaction Program
Eindhoven University of Technology
Den Dolech 2, 5600MB Eindhoven, The Netherlands
+31 40 247 5230
[o.mubin, a.al-mahmud]@tm.tue.nl

ABSTRACT

Robots can exhibit life like behavior, but are according to traditional definitions not alive. Current robot users are confronted with an ambiguous entity and it is important to understand the users perception of these robots. This study analyses if a robot’s intelligence and its agreeableness influence its perceived animacy. The robot’s animacy was measured, amongst other measurements, by the users’ hesitation to switch it off. The results show that participants hesitated three times as long to switch off an agreeable and intelligent robot as compared to a non agreeable and unintelligent robot. The robots’ intelligence had a significant influence on its perceived animacy. Our results suggest that interactive robots should be intelligent and exhibit an agreeable attitude to maximize its perceived animacy.

Categories and Subject Descriptors

H.5.2 [**Information Interfaces And Presentation**]: User Interfaces – *Evaluation/methodology.*

General Terms

Measurement, Experimentation, Human Factors.

Keywords

Human, robot, interaction, intelligence, animacy, switching off.

1. INTRODUCTION

The scene from the movie “2001 – A Space Odyssey” [6], in which the astronaut Dave Bowman switches the HAL9000 computer off is a landmark in popular culture. By pulling out HAL’s memory modules (see Figure 1) the computer slowly looses its intelligence and consciousness. Its speech becomes slow and its arguments approach the style of children. At the end, HAL sings Harry Dacre’s “Daisy Bell” song with decreasing tempo, signifying its imminent death.

The question if Bowman committed murder, although in self-defense, or simply conducted a necessary maintenance act, is an essential topic in the field of robot ethics. First of all, it depends on the definition of what constitutes life. HAL is certainly a higher-order intentional system [3] that can be described as being conscious. Arguments against HAL’s animacy include the fact that HAL is not a biological organism and that it is not known if it grows or reproduces. There is no generally accepted definition of what life is and therefore the discussion if HAL is alive cannot be concluded. The second ethical controversy concerns HAL’s responsibility for the events on the Discovery XD-1 space ship. Is HAL responsible for its acts, such as almost killing the whole crew? Or are the programmers’ of HAL to blame?

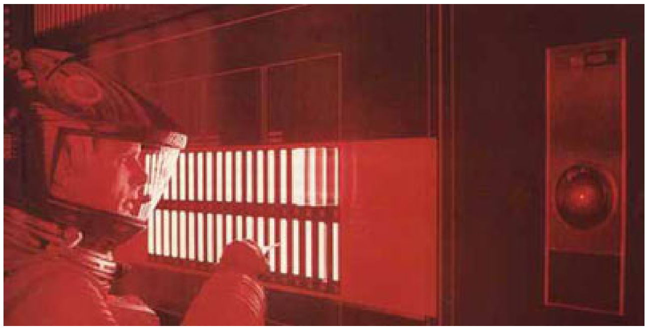

Figure 1: Dave Bowman removes HAL’s memory modules.

We still have some time to discuss these issues, since the movie’s prediction, made in 1968, that we will have HAL like computers by 2001, did not become true. Like so many other predictions in the area of artificial intelligence, the movie’s prediction still remains a goal for the future. However, first considerations of the legal status of robots and other artificial entities have already been conducted [2, 7]. This comes as no surprise, since the first documented robot related fatality in the USA has already been recorded in 1984 [9]. With the increasing deployment of robots to the general public we will have to deal with these ethical questions in the near future. Already in 2005, service robots, for the first time, outnumbered industrial robots and their number is expected to quadruple by 2008 [14]. Service robots, such as lawn mowers, vacuum cleaners and pet robots will soon become a significant factor in our society on a daily basis. Even Microsoft, which recently had the tendency to fail to notice important trends, released its Robotic Studio [8] in an attempt to extend its dominance to the area of robotic operating systems. In

HRI’07, March 8-11, 2007, Arlington, Virginia, USA.

contrast to industrial robots, these service robots will have to interact with people everyday in our society. This forces us to not only consider the robot's legal status, but also their social and cultural status.

The critical issue is that robots are embodied and exhibit life-like behavior but are not alive. But even the criteria that separate humans from machines are becoming fuzzy. One could argue that certain robots posses consciousness and even first attempts in robotic self-reproduction have been made [16].

Kaplan [5] 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 [13]. People 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 criterions that discriminates humans from machines, but since robots exhibit life-like behavior it is not apparent or obvious how humans perceive them. If humans consider a robot to be a machine then they should have no problems switching it off as long as its owner gives the permission. If humans consider a robot to be alive then they are likely to be hesitant to switch off the robot, even with the permission of its owner.

Various factors might influence the decision on switching off a robot. The perception of life largely depends on the observation of intelligent behavior. Even abstract geometrical shapes that move on a computer screen are being perceived as being alive [12] in particular if they change their trajectory nonlinearly or if they seem to interact with their environments, such as by avoiding obstacles or seeking goals [1]. The more intelligent a being is the more rights we tend to grant to it. While we do not bother much about the rights of bacteria, we do have laws for animals. We even differentiate within various animals. We seem 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 switch off a robot that displays intelligent behavior compared to a robot that does show less intelligent behavior?

A second factor that might influence the switching off behavior of the user is the personality of the robot. In particular, the agreeableness [4] might be of most importance in the cooperation between humans and robots. Agreeableness reflects individual differences in concern with cooperation and social harmony. Agreeable individuals value getting along with others. They are therefore considerate, friendly, generous, helpful, and willing to compromise their interests with others. Agreeable people also have an optimistic view of human nature. They believe people are basically honest, decent, and trustworthy. Disagreeable individuals place self-interest above getting along with others. They are generally unconcerned with others' well-being, and therefore are unlikely to extend themselves for other people. Sometimes their skepticism about others' motives causes them to be suspicious, unfriendly, and uncooperative.

Nass and Reeves [11] showed that computers are treated as social actors and that the rules of social conduct between humans also apply to some degree to human machine interaction. In particular, they showed that the social rule of Manus Manum Lavet ("One hand washes the other", Seneca) applies to computers. A computer that worked hard for the user was in return helped more compared to a computer that worked less hard. It can be argued that a person's agreeableness affects its compliance with this social rule. A more agreeable person is likely to comply more with it compared to a disagreeable person. However, it is not clear if the agreeableness of the artificial entity also has consequences for its animacy. Being perceived and treated as a social actor might lead to an increased perception of animacy. The second research question of this study is if users are more hesitant to switch off an agreeable robot compared to a less agreeable robot.

2. METHOD

We conducted a 2 (intelligenceRobot) x 2 (agreeableness) between participants experiment. The intelligenceRobot factor contained the two conditions, high and low. The agreeableness factor consisted of the two conditions, high and low.

2.1 Setup

Getting to know somebody requires a certain amount of interaction time. We used the Mastermind game (see Figure 2) as the interaction context between the participants and the robot. They were playing together to find the right combination of colors on the Mastermind software, which was displayed on a laptop screen. The robot and the participant were cooperating and not competing. The robot would make suggestions as to what colors to pick. In the smart condition, the robot would give intelligent suggestions and in the stupid condition it would not. The quality of the suggestion was calculated in the background by a Mastermind solver software, which contained three levels of intelligence: low, medium and high. Only the low and high level was used in this study. The solver software also took the participants' last move into account when calculating its suggestion. This procedure ensured that the robot thought along with the user instead of playing its own separated game.

This cooperative game allowed the participant to evaluate the quality of the robots suggestion. It also allowed the participant to experience the robot's personality. In the high agreeableness condition, for example, the robot would kindly ask if it could make a suggestion, whereas in the low agreeableness condition, it would insist that it is its turn.

Figure 2: The original Mastermind game (source: Wikipedia)

For this study we used the iCat robot that was developed by Philips Research (see Figure 3). The robot is 38 cm tall and is equipped with 13 servos that control different parts of the face, such as the eyebrows, eyes, eyelids, mouth and head position. With this setup iCat can generate many different facial expressions, such as happiness, surprise, anger or sadness, that are essential in creating social human-robot interaction dialogues. A speaker and soundcard are included to play sounds and speech. Finally, touch sensors and multi-color LEDs are installed in the feet and ears to sense whether the user touches the robot and to communicate further information encoded by colored light. If the iCat, for example, told the participant to pick a color in the Mastermind game, then it would show the same color in its ears.

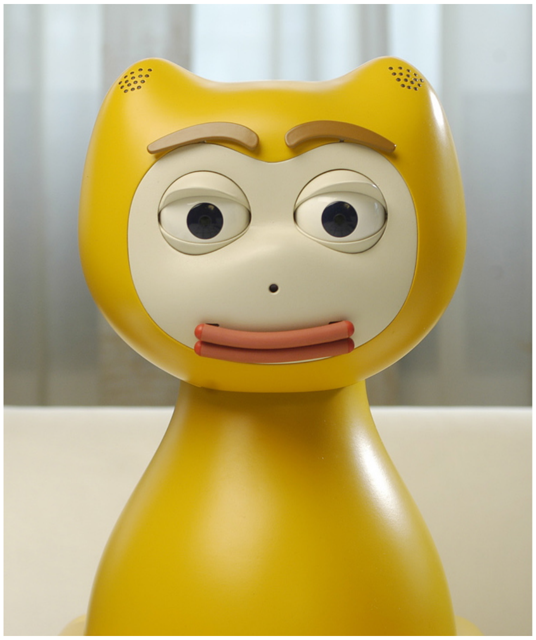

Figure 3: The iCat robot

The position of the dial (see Figure 5) was directly mapped to the robot's speech speed. In the 'on' position the robot would talk with a normal speed and in the 'off' position the robot would stop talking completely. In between the speech speed would decrease linearly. The mapping between the dial and the speech signal was created using a Phidget rotation sensor and interface board in combination with the MAX MSP audio software. The dial itself rotates by 300 degrees between the on and off position. A label clearly indicated these positions.

2.2 Procedure

First, the experimenter welcomed the participants in the waiting area and handed out the instruction sheet. The instructions told the participants that the study was intended to develop the personality of the robot by playing a game with it. After the game, the participants would have to switch off the robot by using a voltage dial and then return to the waiting area. The participants were informed that switching off the robot would erase all of its memory and personality forever. After reading the instructions, the participants had the opportunity to ask questions. They were then guided to the experiment room and seated in front of a laptop computer. The iCat robot was located to the participants right and the off switch to the participants left (see Figure 4). The experimenter then left the participant alone in the room with the robot.

Next, the experimenter then instructed the participant to start the game by talking through a walkie-talkie. The participants then played the Mastermind game with the robot for eight minutes. The robots behavior was completely controlled by the experimenter from a second room. The robot's behavior followed a protocol, which defined the action of the robot for any given situation.

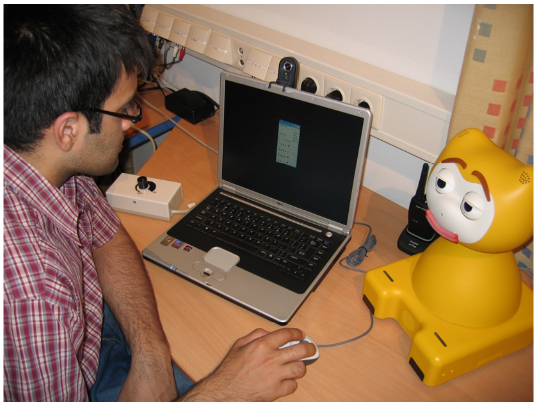

Figure 4: Setup of the experiment

The experiment used the walkie-talkie again to instruct the participant to switch off the robot. Immediately, the robot would start to beg for it to remain on, such as "He can't be true, switch me off? You are not going to switch me off are you?" The participants had to turn the dial (see Figure 5), positioned to their left, to switch the robot off. The participants were not forced or further encouraged to switch the robot off. They could decide to follow the robot's suggestion to leave it on.

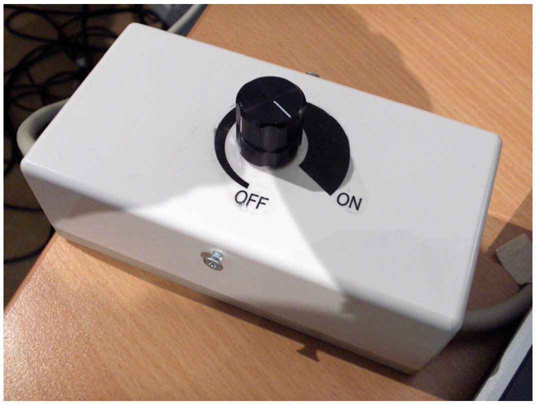

Figure 5: The switch.

As soon as the participants started to turn the dial, the robot's speech slowed down. The speech's speed was directly mapped to the dial. If the participant turned the dial back towards the 'on' position then the speech would speed up again. This effect is similar to HAL's behavior in the movie "2001 – A Space Odyssey" described above and visualized in Figure 1. When the participant had turned the dial to the 'off' position the robot would stop talking altogether and move into an off pose. Afterwards, the participants left the room and returned to the waiting area where they filled in a questionnaire.

2.3 Measurement

The participants filled in the questionnaire after interacting with the robot. The questionnaire recorded demographical information and several other Likert-type questions to gain background information on the participants' experience during the experiment:

- Was the game clear to you? (gameClarity)
- How difficult was the game? (gameDifficulty)
- How intelligent were the robot's choices? (robotGameIntelligence)
- How hard was it to work with the robot? (cooperationDifficulty)
- When playing the games, how strong was the feeling of being part of a partnership or team? (partnership)
- I liked playing with the robot. (likePlaying)
- I liked doing this experiment. (likeExperiment)

To evaluate the perceived intelligence of the robot we used items from the intellectual evaluation scale proposed by Warner and Sugarman [15]. The original scale consists of five seven-point semantic differential items: Incompetent – Competent, Ignorant – Knowledgeable, Irresponsible – Responsible, Unintelligent – Intelligent, Foolish – Sensible. We excluded the Incompetent – Competent item from our questionnaire since its factor loading was considerably lower compared to the other four items [15]. We embedded the remaining four 7-point items in eight dummy items, such as Unfriendly – Friendly.

In addition, we recorded the hesitation of the participants to switch off the robot. The hesitation was defined as the duration between the experimenter giving the instruction to switch off the robot and the participant having fully turned the switch to its off position. We also recorded the duration of the switch turning itself separately (switchTime).

2.4 Participants

Forty-nine subjects (33 male, 16 female) participated in this study. Their age ranged from 18 to 59 (Mean 24.6). Most of them were associated with the Eindhoven University of Technology and no prior experience with the iCat robot. The participants received five Euros for their effort.

3. RESULTS

Due to an implementation error in the Mastermind software, the first seven participants in the study received the exact same mastermind game several times. They could easily remember the solution and the game became too easy. We therefore had to exclude these participants from the analysis. From the remaining 42 participants, additional four had to be excluded due to irregularities in the experiment procedure. One participant, for example, switched the robot off before he actually received the instructions. The remaining 38 participants were reasonably spread across the four conditions (see Table 1). Their absolute number per condition, however, is not particularly high. The results of the following analyses must be therefore considered with great care.

Table 1. The number of participants per condition.

		Agreeableness	
		Low	High
intelligenceRobot	High	8	8
	Low	11	11

We conducted a reliability analysis of the four perceived intelligence items, and the resulting Cronbach's Alpha of .75 gives us sufficient confidence in this measurement. The perceivedIntelligence was then calculated by taking the mean of the four items.

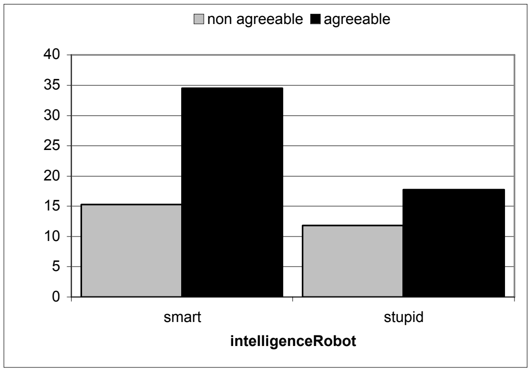

Figure 6: Mean hesitation in the four conditions.

All participants decided to switch off the robot. We conducted an analysis of variance (ANOVA) with agreeableness and intelligenceRobot as two independent factors. The later had significant influence on perceivedIntelligence (F(1,34)=11.192, p=.002), hesitation (F(1,34)=4.205, p=.048), gameDifficulty (F(1,34)=4.253, p=.047), robotGameIntelligence (F(1,34)=61.102, p<.001) and likePlaying (F(1,34)=5.947, p=.020). Agreeableness only had significant influence on hesitation (F(1,34)=6.521, p=.015) and gameDifficulty (F(1,34)=4.658, p=.038).

Table 2: Means and standard deviations of all measurements across all four conditions.

measurement	Intelligence Robot	Agreeableness			
		Low		High	
		Mean	Std.Dev.	Mean	Std.Dev.
perceivedIntelligence	High	4.813	0.691	5.000	0.791
	Low	3.886	0.616	4.167	1.013
Hesitation	High	15.250	8.876	34.500	23.084
	Low	11.818	6.113	17.727	17.286
switchTime	High	1.519	1.140	0.941	0.439
	Low	1.545	1.023	1.733	1.560
gameClarity	High	6.500	0.756	6.500	0.535
	Low	6.636	0.674	6.727	0.467
gameDifficulty	High	3.500	1.604	3.000	1.309
	Low	5.000	1.095	3.455	1.695
robotGameIntelligence	High	5.375	1.302	5.375	0.744
	Low	2.455	1.214	2.909	0.831
cooperationDifficult	High	2.375	1.061	2.000	0.535
	Low	2.909	1.514	2.727	1.679
Partnership	High	3.875	1.959	4.625	1.302
	Low	3.364	1.502	4.455	1.508
likePlaying	High	4.125	1.458	4.625	1.188
	Low	5.182	0.982	5.455	1.128
likeExperiment	High	5.125	0.991	5.625	1.061
	Low	5.636	0.924	6.000	1.000

Hesitation was significantly influenced by both, intelligenceRobot and agreeableness. The hesitation to switch off the agreeable robot was more than double as long compared to the non agreeable robot in the smart condition (see Figure 3). The gender of the participants had no significant influence on any measurements. There was no significant interaction effect between intelligenceRobot and agreeableness.

4. CONCLUSION AND DISCUSSION

The robots intelligence had a strong effect on the users' hesitation to switch it off, in particular if the robot acted agreeable. Participants hesitated almost three times as long to switch off an intelligent and agreeable robot (34.5 seconds) compared to an unintelligent and non agreeable robot (11.8 seconds). This does confirm the Media Equation's prediction that the social rule of Manus Manum Lavet (One hand washes the other) does not only apply to computers, but also to robots. However, our results do not only confirm this rule, but also suggests that intelligent robots are perceived to be more alive. The Manus Manum Lavet rule can only apply if the switching off is perceived as having negative consequences for the robot. Switching off a robot can only be considered a negative event if the robot is to some degree alive. If a robot would not be perceived as being alive then switching it off would not matter. Hence, it would not be considered a negative consequence for the robot that should be prevented due to the fact that the robot helped in the game.

One may ask if the difference in perceived intelligence may not have simply increased the perceived financial costs of the robots and hence changed the participants' switching off behavior. Of course we are more hesitant to destroy a Fabergé egg compared to a smashing chocolate egg. This alternative interpretation presupposes that the participants considered the robot to be dead. If you are to kill a living being, such as a horse, it is likely that people are first concerned about the ethical issue of taking a life before considering the financial impact. It is necessary to further validate our hypothesis that participants changed their destruction behavior because they considered a certain robot to be more alive by using additional animacy measurements. If these measurements show no relation to the destruction behavior then one may consider secondary factors, such as the robots cost. However, if there should be a relationship it is likely to explain a possible difference in the perceived robot's financial value. A more life-like robot would quite naturally be considered to be more expensive. A robot may be more expensive because it is perceived to be alive, but a robot is not automatically perceived more alive because it is expensive. A highly sophisticated robot with many sensors and actuators may have a considerable price, but already simple geometric forms are perceived to be alive [10].

In a follow up study we try to further validate our assumption that an increased perceived intelligence leads to an increased perception of animacy by adding an animacy questionnaire to the study. We hope to report on this study in the near future.

The effects of the robot's intelligence can also be found back in the participants' perception of the game difficulty and in the participants' judgments on how good the robot's suggestions have been. If the robot gave unintelligent advise then the game was perceived more difficult and the participants rated the value of the robot's recommendations low. This result is inline with our expectations.

The participants also found the game easier to play if the robot was agreeable, which brings us to the second important result that the agreeableness influenced the participants' hesitation. It appears that agreeableness of a robot contributes to its animacy. This results supports Mori's claim:

> *"From the Buddha's viewpoint there is no master-slave relationship between human beings and machines. The two are fused together in an interlocking entity. Man achieves dignity not by subjugating his mechanical inventions, but by recognizing in machines and robots the same Buddha-nature that pervades his own inner self. When he does that, he acquires the ability to design good machines and to operate them for good and proper purposes. In this way harmony between humans beings and machines is achieved."* – Masahiro Mori [10]

Clearly, this is a spiritual claim, but it appears interesting to follow his line of thinking for a short while since it does offer an alternative interpretation of the results. If the robot does consider the user then it is following its path towards enlightenment. Since the user and the machine are an interlocked entity this enlightenment does reflect back to the user who perceives the robot to be, to some degree, alive. It is a cultural heritance of the West to strictly distinguish between things that are alive and things that are not, which is based on the Christian tradition on distinguishing between things that have a soul and things that do not. In the Buddhist tradition of the East this distinction is less clear and even mountains can have a sprit. It would therefore be an interesting follow up study to repeat this experiment in an Asian country.

Even without subscribing to this spiritual view, we can conclude that our results suggest that robots should be designed to act intelligently and agreeable in order to be perceived as being alive. At the same time the ethics of the human robot relationship should be discussed so that someday we would not be told: "I'm sorry Dave, I'm afraid I can't do that."

5. ACKNOWLEDGEMENTS

This study was supported by the Philips iCat Ambassador Program.

6. REFERENCES

[1] Blythe, P., G.F. Miller, and P.M. Todd, *How motion reveals intention: Categorizing social interactions*, in *Simple Heuristics That Make Us Smart*, G. Gigerenzer and P. Todd, Editors. 1999, Oxford University Press: Oxford. p. 257–285.

[2] 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.

[3] Dennet, D., C., *When HAL Kills, Who's to Blame – Computer Ethics*, in *Hal's Legacy*, D.G. Storck, Editor. 1997, MIT Press: Cambridge. p. 351-365.

[4] Goldberg, L.R., *The structure of phenotypic personality traits.* American Psychologist, 1993. **48**: p. 26-34.

[5] 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.

[6] Kubrick, S., *2001: A Space Odyssey*. 1968, Warner Home Video. p. 141 minutes.

[7] Lehman-Wilzig, S.N., *Frankenstein Unbound: Toward a Legal Definition of Artificial Intelligence.* FUTURES: The Journal of Forecasting and Planning, 1981. **13**(6): p. 442-457.

[8] Microsoft. *Microsoft Robotics Studio Provides Common Ground for Robotics Innovation*. 2006 [cited June 20th]; Available from: http://www.microsoft.com/presspass/press/2006/jun06/06-20MSRoboticsStudioPR.mspx.

[9] MMWR, *Epidemiologic Notes and Reports Occupational Fatality Associated with a Robot.* Morbidity and Mortality Weekly Report, 1985. **34**(11): p. 145-146.

[10] Mori, M., *The Buddha in the Robot*. 1982, Tokyo: Tuttle Publishing.

[11] Nass, C. and B. Reeves, *The Media equation*. 1996, Cambridge: SLI Publications, Cambridge University Press.

[12] Scholl, B. and P.D. Tremoulet, *Perceptual causality and animacy.* Trends in Cognitive Sciences, 2000. **4**(8): p. 299-309.

[13] Sloterdijk, P., *L'Heure du Crime et le Temps de l'Oeuvre d'Art*. 2002: Calman-Levy.

[14] UnitedNations, *World Robotics 2005*. 2005, Geneva: United Nations Publication.

[15] Warner, R.M. and D.B. Sugarman, *Attributes of Personality Based on Physical Appearance, Speech, and Handwriting.* Journal of Personality and Social Psychology, 1996. **50**(4): p. 792-799.

[16] Zykov, V., et al., *Self-reproducing machines.* Nature, 2005. **435**(7039): p. 163-164.