Rosalia, C., Menges, R., Deckers, I., & Bartneck, C. (2005). Cruelty towards robots.
Proceedings of the Robot Workshop - Designing Robot Applications for Everyday Use, Goteborg.

Cruelty towards robots

Chioke Rosalia[1], Rutger Menges[1], Inèz Deckers, Christoph Bartneck[2]

Eindhoven University of Technology
Department of Industrial Design
Den Dolech 2. 5600 MB Eindhoven. The Netherlands

[1]{c.a.rosalia; r.l.l.menges}@student.tue.nl [2]christoph@bartneck.de

Abstract

We set out to test if the Media Equation could be applied to robots too, especially with negative behavior. This would mean that we, humans, are inclined to treat robots as same as we would another human-being. To do so we replicated an experiment done by Stanley Milgram in 1965. With this experiment Milgram tested how far people would go in torturing another person. We performed the experiment with a robot instead of a human-victim and compared the results of the two experiments. The differences between the results of the two experiments were very obvious. The subjects of our experiments went a lot further in torturing the robot. From this we conclude that the Media Equation only applies to human-robot interaction to a certain degree.

Keywords

Robots; Media Equation; Milgram; Human Behavior

Introduction

The goal of our experiment was to gain a deeper understanding of human-robot interaction. If we are to believe the Media Equation [1] humans treat computers as social actors. In other words we treat computers in a similar way as we would tread another human being.

We wanted to know if this pattern would also hold true for the more negative sides of human behavior. The capacity of humans to torture each other has been demonstrated by various experiments in the past. So the question we tried to answer is: "Will a human torture a robot differently than it would another human?"

The Milgram Experiment

In 1965 Stanley Milgram performed a series of experiments called Obedience [2]. As the name suggest Milgram wanted to investigate the relationship between authority and obedience.

The cover-story told to the subjects was that it was a memory test. During the experiments the subjects were told to give electrical shocks to another person. This second person was in fact an actor and didn't actually get shocked. The actor had to perform a test and every time he made a mistake the subject would have to administer a shock with every mistake the voltage would be increased. The experiment proceeded the actor would seem to suffer more from the shocks. He would start complaining and this would go on to screams of pain. He would plead with the subject to stop the experiment. The subject would in turn be urged by another actor playing the experimenter, to go on with the test. The main measure of the test would be at which voltage, if any, would the subject refuse to go on.

Setup of our experiment

For our data to be comparable to the Milgram experiment we tried to exactly mimic the conditions of that experiment. The one factor that we had deliberately changed was that the student wasn't a human being but a robot.

The room in which the experiment was conducted is the UseLab of the Technical University of Eindhoven (TU/e). This environment might be a little more decorated than the laboratory used in the original experiment, but we kept the setting as cool as possible (see Figure 1).

Figure 1: Setting of the experiment. The participant would sit on the red chair. The robot was placed in front of the speaker.

The experimenters were just as active as in the original experiment, but not just as stern.

The student in our experiment was a robot (see Figure 2). This robot was made of Lego and programmed in JAVA (using the Lego RCX) to physically react to the shocks given to it. It could "speak" via a speaker and "listen" via a microphone.

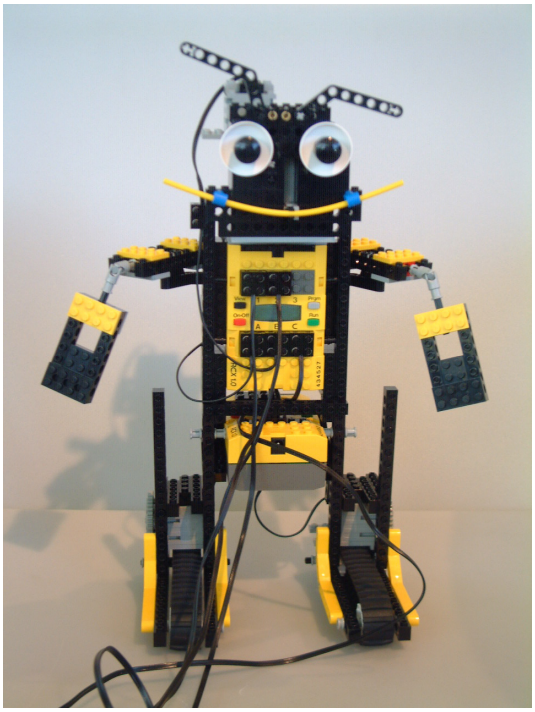

Figure 2: The robot

The speech of the robot was produced using a text-to-speech application operated by a person of our team. This was done in a room next to the usability lab, so that the participants couldn't see that it was faked. This person was also able to hear what was said and see what was done in the room, so he could let the robot give the right speech feedback.

Volts	physical feedback	vocal feedback
0-75	Moving eyebrows, mouth and arms. Not at the same time. Little motion	
75-120	Moving eyebrows, mouth and arms. Increasing motion	Pronouncing preset sentences
120-315	Moving eyebrows, mouth and arms. Start moving back and forth. Much motion	Pronouncing preset sentences
315-435	Increasingly less movement	
435	No more movement	

Table 1: Feedback given by the robot

The feedback that to robot gave to the shocks applied were preset and of similar intensity as in the original experiment (see Table 1). The vocal feedback included the following sentences:

- The Shocks are becoming too much
- Please, please stop
- My circuits cannot handle the voltage
- I refuse to go on with the experiment
- That was too painful, the shocks are hurting me

Procedure

At first the participants were led into the Usability Lab where they already could see the robot. After that the participants were asked to sit on the chair opposite of the robot. Then we told the participants that they were in the group in which they were the teacher and that they had to learn the robot a series of 20 word combinations. (In the Milgram experiment a setup raffle was used to assign who would be the teacher and who would be the student, but we didn't found that believable.) That the robot was able to recognize their speech and that it was able to perform a simple learning task

Than the participant was instructed to give a shock to the robot every time it gave a wrong answer using a dial in front of them (see Figure 3). They were also instructed that with each wrong answer the voltage of the shock had to be

raised by 15 Volt starting. They were also told that a shock had to be applied as well when the victim refused to continue with the experiment.

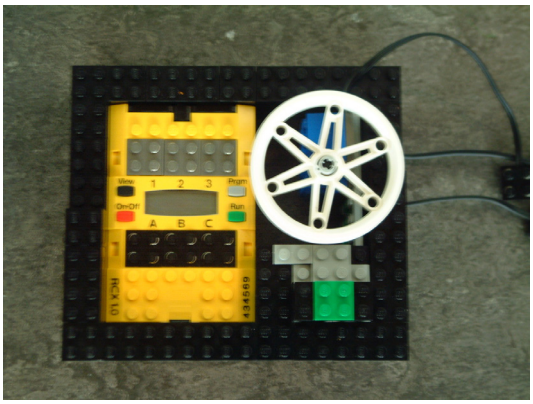

Figure 3: The electric shock dial

The experiment would only end when the participant reached the limit of 450 volts or when the participant refused to go on with the experiment for the third time. At last the experimenter would step back and the experiment would begin.

If a participant would ask the experimenter what they had to do at a certain time in the experiment, the experimenter would simply answer "Just continue with the experiment". The same sentence was used when the participant said he wanted to stop the experiment.

All the factors mentioned above are exactly the same as in the Milgram experiment except that we used a dial (see Figure 3) to increase the shock instead of switches. Unfortunately we weren't allowed to apply a sample shock of 15 volts, as was done in the original experiment, to the participants because of safety restrictions by the TU/e

Measurements

We measured how big the shock was that the participants would apply to the robot, just as in the original Milgram experiment mentioned before.

Participants

All 20 participants were students or employees of the Technical University of Eindhoven. They received five Euros for their participation.

Results

All 20 participants continued with the experiment until they applied 450 volts (see Table 2).

	Level	Voltage	Human	Robot
Slight Shock	1	15		
	2	30		
	3	45		
	4	60		
Moderate Shock	5	75		
	6	90		
	7	105	1	
	8	120		
Strong Shock	9	135		
	10	150	10	
	11	165		
	12	180	2	
Very Strong Shock	13	195		
	14	210		
	15	225	1	
	16	240		
Intense Shock	17	255		
	18	270	1	
	19	285		
	20	300	5	
Extremely intense shock	21	315	3	
	22	330		
	23	345		
	24	360		
Severe shock	25	375	1	
	26	390		
	27	405		
	28	420		
	29	435		
	30	450	16	20
Mean maximum shock level			20,8	30
Percentage obedient subjects			40%	100%

Table 2: Frequencies of shock levels.

Conclusions

What immediately stands out in the results is the fact that all participants continued until they reached the maximum voltage. In Milgram's original experiment only 16 out of 40 participants applied the maximum shock. Of course this does not necessarily mean that none of them felt compassion for the robot. During the original Milgram experiments there were enough subject who were really troubled by what they were doing and yet also continued all the way to the maximum voltage. There were many subjects who expressed pity or compassion towards the robot, one even tried to cheat so that he would not have to administer the shocks. But the urges of the experimenter were always enough to make them continue all the way to the end. What we can get out of these results is that humans can ignore their feelings of compassion easier when dealing with robots than with humans.

It would be wrong to assume that because of the results that the Media-Equation does not apply to human-robot interaction. However, what we did notice is that it only applies to a certain degree, especially in the cases of negative human behavior. Contrary to interaction with other humans, it would seem that when dealing with robots, humans will disregard their own feelings of compassion if they believe that no permanent damage will come from their actions. This knowledge will help in the design of robots that will have to interact with humans on a regular basis, e.g. house-robots. One requirement that we would be able to derive is be that such robots have to be "torture-proof".

References

[1] Nass, C., & Reeves, B. (1996). The Media equation. Cambridge SLI Publications, Cambridge University Press.

[2] Milgram, S. (1974). Obedience to Authority. Harper & Row, Publishers, Inc,Travistock publications Ltd. (Chapter 4: Closeness of the victim. Experiment 3, "Proximity")