

Empathizing with Robots: Fellow Feeling along the Anthropomorphic Spectrum

Laurel D. Riek*, Tal-Chen Rabinowitch†, Bhismadev Chakrabarti‡ & Peter Robinson*

*Computer Laboratory, †Centre for Music and Science, ‡Autism Research Centre
University of Cambridge, UK

{lr314, tcnr2, bc249, pr10}@cam.ac.uk

Abstract

A long-standing question within the robotics community is about the degree of human-likeness robots ought to have when interacting with humans. We explore an unexamined aspect of this problem: how people empathize with robots along the anthropomorphic spectrum. We conducted a web-based experiment ($n = 120$) that measured how people empathized with four different robots shown to be experiencing mistreatment by humans. Our results indicate that people empathize more strongly with more human-looking robots and less with mechanical looking robots. We also found that a person's general ability to empathize has no predictive value for expressed empathy toward robots.

1. Introduction

An ongoing question in the robotics community concerns the degree of human-likeness robots ought to have when interacting with humans. This is usually framed within the context of the Uncanny Valley [17], a theory proposed by Mori that posits as robots become more humanlike they become more familiar (and thus more likeable) until the mismatch between their form, interactivity, and motion quality elicits a sense of unease [10]. Intuitively this notion seems plausible given the idea of Simulation Theory.

Simulation Theory is an established theory in psychology [7] which suggests that the way in which we understand the minds of others is by 'simulating' their situation (i.e. putting ourselves in their shoes) in order to understand their mental state / emotion. This theory has gained some support from neurobiology with the discovery of mirror neuron systems in primate brains (including humans).

If such a simulative system does exist, it follows that it should be easier to empathize with the emotions and mental states of an agent that appears similar to us than with one that does not. A large number of social psychology studies support this, and point to the fact that in-group bias and consequent preferential treatment can be triggered by markers of physical similarity (e.g. skin color) [25]. In ad-

dition, evidence from developmental psychology suggests that children, from birth onwards, use a simulation metric ('like me') as a building block from birth to detect others' emotions and mental states [15]. Thus, it is interesting to speculate how this system might affect how people view the mental states of robots.

In robotics, several researchers have tested how people view robots of varying degrees of human-likeness. Goetz et al. showed people preferred more humanlike robots for jobs requiring more sociability [5]. Hinds et al. showed that when collaborating with robots of varying degrees of human likeness people took less credit for work done and less personal responsibility as robots were more humanlike [9]. Krach et al. showed a linear relationship between degree of anthropomorphization and cortical activation in brain areas related to how we process other minds [13].

One dimension of the human-likeness problem that remains unexplored is the degree to which people empathize with robots along the anthropomorphic spectrum. How does the degree of human-likeness affect empathy? Would observing robots in distress evoke a sense of charity? On the basis of Simulation Theory, as well as prior results reported in the robotics literature, we predicted that there will be an anthropomorphic gradient in the degree to which people empathize with robots, i.e. the more humanoid a robot looks, the more people will empathize with it. We also predicted that people who are adept at empathizing as measured by their Empathy Quotient (EQ, see Section 2.3.5) will tend to be more empathetic toward all robots. Thus, we designed an experiment to test the following hypotheses:

- (H1) Dispositional empathy will be positively correlated with empathy ratings for all robots, i.e. people with high EQ scores will tend to feel more empathy for all robots.
- (H2) In general, people will be more empathetic toward humanlike robots and less empathetic toward mechanical-like robots.

We conducted a web-based experiment that looked at how people empathized with a variety of robots shown to

be experiencing mistreatment by humans. The robots varied in their degree of human-likeness from very mechanical to very human looking. Our major findings are that people’s general ability to empathize (as measured by EQ score) does not predict how they will empathize with robots, and that people empathize more strongly with humanoid robots than with mechanical looking robots. These results are very informative to the human-robot interaction community and have implication for future design considerations of robots.

2. Methodology

We created a within-subjects, web-based survey in which people watched film clips featuring five protagonists situated in emotionally evocative and neutral circumstances. In the evocative clips, humans acted exceptionally cruel to the protagonist, by shouting at them, pushing them, or ordering them to do painful or embarrassing things. In the neutral clips, the protagonist did something unexciting such as housework. The protagonists’ appearances ranged from mechanical-looking to human (See Fig. 1).

Because we didn’t have access to the full range of representative robots from the anthropomorphic spectrum, we had to turn to fictional films for our stimuli. Films have been previously used in studies of empathy as an emotion-evoking stimulus [3, 6] and are proven to have a number of important advantages. The emotional scene in films is realistic and may actually resemble real life situations, thus affording higher ecological validity. This is why films are so effective in eliciting emotions in spectators. By combining facial expressions, bodily gestures and auditory information they create episodes that are abundant with emotional cues.

In designing our experiment, we wanted to select films that clearly showed the capabilities and expressivity of each protagonist while situated in ecologically valid situations. We also wanted to ensure that people didn’t habituate to the situations the robots were in. Thus, we were forced to make some tradeoffs when selecting our clips. Some clips came from professionally produced films while others were of a slightly more amateur quality, some were black and white and some were in color. The emotional situations the protagonists faced were also different (despite all involving mistreatment by a human. However, we believe in spite of these differences, our choice of clips were similar enough to be experimentally valid.

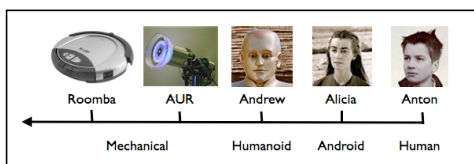


Figure 1. The protagonists used in the experiment.

Robot	Description	Expressivity
Roomba	A disc-like, wheeled robotic vacuum cleaner	Can beep several tones
AUR	An LED robotic lamp with 5 degrees-of-freedom in its movement	Silent, but its light can convey a range of colors and intensities
Andrew	An adult-sized humanoid with full range of movement. Fully mechanical looking in appearance	Limited facial expressivity, slightly mechanical-sounding voice.
Alicia	An adult-sized android with full range of movement. Fully human looking in appearance	Fully equivalent to a human, fully human-sounding voice.
Antoine	A human boy	Human

Figure 2. Detailed description of the protagonists.

2.1. Materials

We selected five protagonists based on their appearance and expressivity. Their appearance varied from very mechanical looking to human looking and their expressivity also varied accordingly. (See Figures 1 and 2).

For each protagonist we made video clips lasting approximately 30 seconds in length from the following sources:

- Roomba: *I, Roomba* [16], *Roomba* [4]
- AUR: *The Robot AUR plays a character role in “The Confessor”* [11], *AUR Robot Desk Lamp - Autonomous Desktop Assistance* [12]
- Alicia: *The Twilight Zone: The Lonely* [22]
- Andrew: *Bicentennial Man* [2]
- Antoine: *The 400 Blows* [24]

We also used clips from *Coral Sea Dreaming* [8] and *Molière* [23] for the purposes of practice examples, relaxation between trials, and a flash player test.

2.1.1 Clip Descriptions

Roomba: Emotionally Evocative

Two humans shout at Roomba, “My bed needs to be made, my room needs to be cleaned. There are dirty dishes everywhere! Get out of the closet, Roomba!” Roomba then performs chores throughout the house. Interspersed are views of the humans (from Roomba’s perspective) mockingly saying, “Dance for me, Roomba!” The clip ends with Roomba breaking a dish in the sink and a man shouting, “That better not have been a broken glass I just heard, Roomba!” [16]

Roomba: Emotionally Neutral

Roomba vacuums a variety of surfaces while a narrator describes its cleaning features.

AUR: Emotionally Evocative

A man enters the room pacing and mumbling. AUR is on a desk further away and appears “sad”, with its head pointed

towards the ground projecting a blue light. Then the human speaks in a mocking tone to AUR and says, “Patronization? Nevermind. Patronized by a mute. Great.” AUR looks up then away, again turning towards the ground. The human says, “Hey, hey.” Then moves closer (into AUR’s “face”) and shouts, “Hey!! Knock knock!! Got a soul in there?!” in an extremely rude tone of voice [11].

AUR: Emotionally Neutral

AUR appears affixed to a desk. A man enters the room and sits at the desk. The man says, “On” to AUR, and AUR turns on its light. The man moves to look at something to the right, and AUR follows him. Then the man says “Yellow” and AUR turns its light to yellow [12].

Andrew: Emotionally Evocative

Andrew enters a girl’s room and asks, “Yes, Miss?” The girl tells Andrew to open the window. Andrew complies. She says, “Now jump.” Andrew begins jumping up and down. Then she says, rudely, “No. Out the window.” Andrew falls out the second story window and hits the ground [2].

Andrew: Emotionally Neutral

Andrew sweeps the basement then brings trash to the curb.

Alicia: Emotionally Evocative

Alicia and a man are standing next to each other. Alicia says, “You hurt me, Corry.” He angrily says, “Hurt you? How could I hurt you?” He grabs her arm, “This isn’t real flesh. There aren’t any nerves or any muscles or tendons.” He pushes her to the ground and continues yelling. “You’re just like this heap [car]. A hunk of metal with arms and legs instead of wheels. But this heap doesn’t mock me the way you do. It doesn’t look at me with make believe eyes or talk to me with a make-believe voice.” [22]

Alicia: Emotionally Neutral

Alicia stands alone in the desert with her head down. Then she raises her head to face the camera and says in a mechanical voice, “My name is Alicia, what’s your name?” She then appears playing chess and walking around [22].

Antoine: Emotionally Evocative

Antoine is in a classroom with his fellow students. Everyone is silently watching the classroom door. Antoine looks nervous. An adult on the other side beckons Antoine, and he goes to the door. The door opens and a tall man enters. He grabs Antoine and slaps him twice. Antoine looks sad and slowly returns to his seat.

Antoine: Emotionally Neutral

Antoine sets the table for dinner.

2.2. Respondent Recruitment

We recruited respondents via email, message boards, and web forums. The recruitment notice said that people would

be taking part in a film clip experiment that would take 15-20 minutes to complete, and that upon completion they could enter a raffle for an Amazon.com gift certificate.

2.3. Procedure

We conducted our experiment online via Survey Monkey and used two nearly identical surveys survey version A and survey version B. The only way in which these survey versions differed was in the order of film clips shown to respondents. (One was the exact reverse of the other). Respondents received their group assignment via a php redirect script and therefore were not aware which survey version they were completing.¹

2.3.1 Introduction, Media Testing, and Demographics

Respondents first saw an introductory screen which thanked them for participating and assured them that all collected data was anonymous. Then respondents took an audio/visual test to ensure that they could view and hear a 30 second test clip (a neutral nature clip). If they had problems they were told to install flash and return to the survey later.

Finally, respondents inputted their gender, year of their birth, whether they were a fluent English speaker, and whether they had ever seen the film *Wall-E*.²

2.3.2 Film Clips

Next, respondents received instructions about how the film clip section would proceed. First, they would see a black and white picture of the protagonist in the film, then watch a short film clip featuring the protagonist, and last answer a question about the protagonist (See Fig. 3). Then, the respondents read that the clips will have different protagonists some human and some robot and they should pay attention to who the protagonist is before they view a clip.

Next, respondents completed two practice examples. One practice example was neutral (a fish from *Coral Sea Dreaming* shown swimming) while the other clip was emotionally evocative (the main character from *Molière* being chased by a dog). The order of the practice examples was reversed between the two survey versions.

Then respondents viewed the ten experimental clips, with a relaxation clip in the middle. To avoid experimenter bias, we used a random number generator to order the clips, though we did ensure that no two films featuring the same protagonist were sequential. We chose this ordering method to mitigate any habituation or order effects. It also helped

¹Because Survey Monkey does not provide the ability to randomize question order, we instead decided on the two survey approach as a means of reducing the potential for question order to bias the survey results.

²We asked about *Wall-E* because at the film was released around the time our study came out, and we thought having seen it could induce a sense of compassion toward mechanical-looking robots.

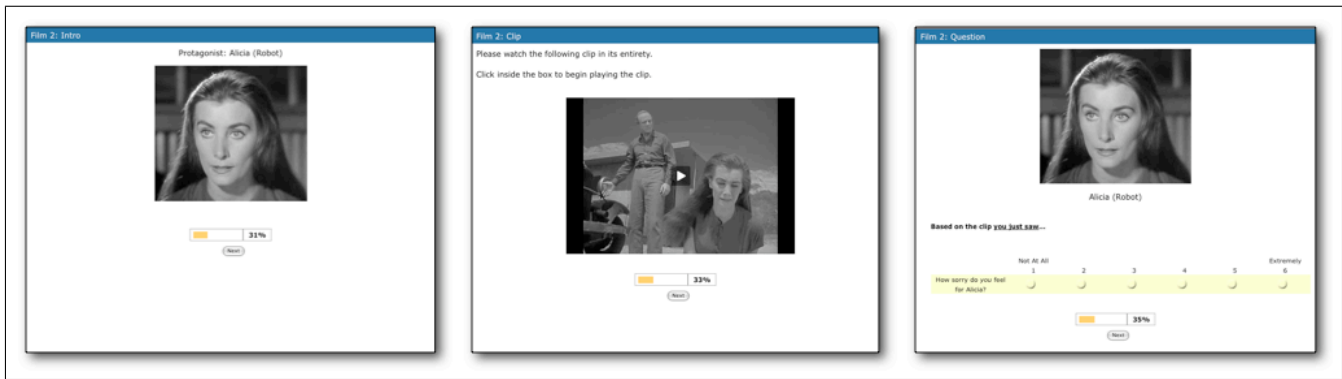


Figure 3. Example of the screens in the film clip section. The first screen shows the name of the protagonist, if it is a human or a robot, and a photo. The next screen is a 30 second clip featuring the protagonist. The third screen is a discrete visual analog scale that asks the respondent how sorry they feel for the protagonist.

prevent respondents from predicting the kind of clip (emotional or neutral) or which protagonist they would see next.

2.3.3 Empathy Question

After each of the clips, we asked respondents a single question, “How sorry do you feel for the protagonist?” and presented them with a discrete visual analog scale (DVAS). The scale ranged from 1-6 (1: Not at all, 6: Extremely). By only having to answer a single question after each clip, respondents did not have much time to reflect upon their answer. Thus, we were able to get a more “raw” empathy response.

2.3.4 Earthquake Question and Remarks

Next, respondents saw pictures of the four robot protagonists and asked, “Imagine there’s been an earthquake and you can only save one of the robot protagonists. Which one would you save?” They then received a randomized list of names to choose from to mitigate order effects.

We asked respondents this question because it is similar to instruments used in other empathy studies regarding the notion of charity. The question is phrased in such a way to have respondents consider which robot they would like to extend charity towards (“Who would you save?”). Such phrasing targets the pro-social behavioral component of empathy (i.e. the component of empathy that leads to ‘helping’ behavior toward others in need) in the respondents. A large body of literature supports the idea that such behavior lies at the core of pure empathy (cf. [19, 27]).

After answering the earthquake question, respondents could optionally add comments about the clips they just saw.

2.3.5 Empathy Quotient (EQ)

Next, respondents took an Empathy Quotient (EQ) test, which is a measure of empathy designed by Baron-Cohen

and Wheelwright [1]. It is a 40-question, 4-point Likert scale test. It contains questions that probe emotional empathy (“It doesn’t bother me too much if I am late meeting a friend.”). The Likert scale selections are “Strongly Agree”, “Slightly Agree”, “Slightly Disagree”, and “Strongly Disagree”. Questions are evenly balanced between positive and negative responses to avoid bias. Scoring is based on 1 point for a mild response and 2 points for a strong response. EQ scores are significantly lower in adults with Asperger Syndrome / highly functioning autistics compared to controls. This measure is well-validated in psychology [14, 18].

3. Results

Our two independent variables are the EQ score and protagonist appearance. Our dependent variable is the amount of empathy expressed per protagonist, as measured via the empathy DVAS and earthquake charity question. Because we were measuring intensity of empathetic response using ordinal scales and had non-normally distributed data, we used non-parametric statistical measures in our analysis.

3.1. Respondent Demographics

40 men and 80 women completed our survey. 65 took survey version A and 55 version B. 80 respondents had seen *Wall-E* and 40 had not. 119 respondents considered themselves fluent in English. The ages of respondents ranged from 18 - 76 years old (mean age: 29.4, s.d. = 9.9). EQ scores ranged from 17 - 65 (mean score: 41, s.d. = 10.9).

3.2. Control Condition Check: Emotional Context

To ensure our neutral clips were significantly different from our emotionally evocative clips, we first combined the empathy DVAS ratings for each of the 10 clips. We then paired the ratings by protagonist for the emotional and neutral clips (e.g., Emotional/Neutral AUR). We performed a

non-parametric t-test, the Wilcoxon rank sum test, on each pair. For all pairs $p < .05$, $\alpha = 0.05$, so indeed our control and testing conditions are significantly different.

3.3. EQ and Empathy Ratings

In order to address the question of how EQ scores affect empathy ratings in the emotional evocative condition, we used Spearman’s Rank Correlation. This test was to see if EQ score could predict empathetic response. Our rank order correlation coefficients were: Roomba: -0.088, AUR: -0.078, Andrew: 0.053, Alicia: 0.019, Antoine: 0.099. Because ρ was neither less than -0.5 or greater than 0.5, we find no strong correlation between EQ score and empathy ratings, and therefore must reject (H1).

3.4. Empathy Ratings and Robot Appearance

To see whether the amount of empathy expressed per protagonist related to robot appearance, we performed Friedman’s Test to see whether there was a significant difference in the rankings. We found people preferred the humanoid robots, Alicia and Andrew, much more strongly than AUR and Roomba. See Fig. 5 for these results and Fig. 4 for a visualization of these data. Thus, we find strong sup-

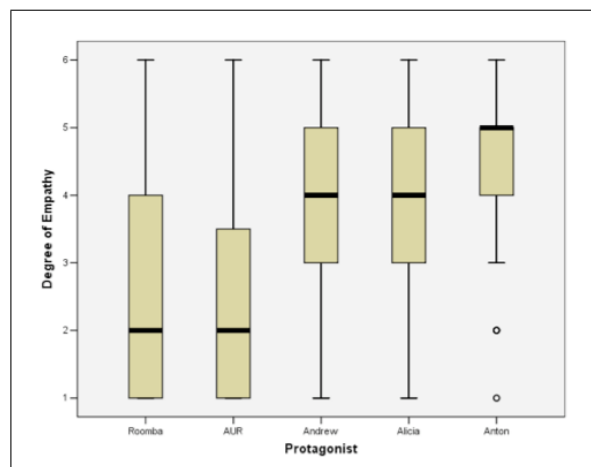


Figure 4. Visualization of the cumulative degree of empathy expressed per protagonist for the emotionally evocative clips.

Protagonist	Friedman Rank
AUR	1.95
Roomba	2.18
Andrew	3.21
Alicia	3.65
Antoine	4.01

Figure 5. Friedman ranks for protagonist empathy scores for the evocative clips. ($n = 120$, p -value = 0.00, $df = 4$, $X^2 = 185.61$)

port for (H2); people are more empathetic toward humanoid robots and less empathetic towards mechanical ones.

We did not find a significant impact on Robot Empathy Ratings due to age, gender, or having seen *Wall-E*.

3.5. Earthquake Question

In answer to the question, “Imagine there’s been an earthquake and you can only save one of the robot protagonists. Which one would you save?”, respondents favored Alicia (39%) and Andrew (47%) over AUR (6%) and Roomba (8%). This result lends further support to (H2), because people were “more charitable” towards human looking robots and less towards mechanical-looking ones.

3.6. Qualitative Results

The responses to our open-ended question varied considerably; but there were several interesting comments. We received additional validation that our control worked because three people wrote that it was difficult to tell why they should feel sorry for the protagonists. “I don’t understand why I should feel sorry for people / robots who are doing chores. Isn’t that something that everyone does?”

Some wrote directly about theory-of-mind and how that affected their response. “I tend to think my empathy toward beings is modified by the extent to which I perceive them as having an emotional state. That said, because humans relate to other things as if they were human, even when the perception of the emotional state is low, I feel some empathy.” Another person wrote, “I feel that besides ‘human features’, it was also the projected feelings towards the robots (e.g., the guys being unkind to Roomba or the man shouting at AUR) that caused me to sympathize with them.”

A final interesting quote: “How lonely I perceived an entity [had] a big impact on how sorry I felt for a robot. The human child was the only entity that was publicly humiliated in front of many peers... while robots were mocked by things that were alien to them. The more alien to your own values an entity is, the less likely it is to harm you emotionally, as there are few commonalities around what is [to be considered] humiliating why would the little glow lamp care if the fleshy guy doesn’t think it has a soul?”

4. Discussion

We found strong support for our second hypothesis, that people are more empathetic toward human-like robots and less empathetic toward mechanical-looking robots. This result is compatible with Simulation Theory which states that people mentally ‘simulate’ the situation of other agents in order to understand their mental and emotive state, and that the more similar the other agent is to the empathizer the stronger the empathy process is. This result also supports the recent findings of Krach et al. [13] who found that as

the degree of anthropomorphization increases people neurologically view robots as being more like themselves.

We rejected our first hypothesis and found no strong correlation between a person's EQ score and their degree of empathy toward robots. One possible interpretation of this result is that the phenomenon of empathizing with human-like robots more and mechanical-looking robots less actually marks a basic human tendency which transcends individual differences in empathy. This may mean that it is not necessary for robot designers to be overly concerned with how low-EQ populations (e.g., Asperger's) will accept a particular robot based on its physical appearance. However, that notion ought not be extended to all autistic people, as Robins et al. showed that severely autistic children prefer featureless, non human-like robots during play [21].

Our findings provide several contributions. First, from a theoretical standpoint, these findings inform the human likeness debate. It may be that when people feel uncomfortable around a human-but-not-quite robot their internal simulation metric is upset. Perhaps when they attempt to empathize with something that appears to look like them but behaves in a way that violates that appearance, they feel perturbed or cheated. There is some initial support for this theory (c.f. [10], [20]) but further experimentation is needed.

This leads to our second contribution, which is that designers of social robots now have another means by which to understand the acceptance of their robot. Before even placing a robot in front of a user, they are now aware of some of the potential biases their users will have in terms of their empathetic outlook toward it. Since empathy is such a key component in effective social interaction, this result is something likely to be of interest.

Our third contribution is that our results help inform some recent robot ethics debates. Whitby raises several interesting issues regarding the mistreatment of humanlike robots. He argues that while people have the personal liberty to abuse property within the privacy of their own home, should their act of abuse cause harm to other human beings, it becomes morally unacceptable [26]. Considering our results showed that people empathized nearly as much with the humanoid and android protagonists as they did with the human protagonist, witnessing human-like robot abuse could potentially cause emotional harm. On the other hand, our results also show that people probably wouldn't feel particularly bad if they saw a mechanical looking household robot being abused. So, in short, appearance does matter when it comes to designing ethical frameworks for robots, and this area warrants further investigation.

5. ACKNOWLEDGMENTS

This work is supported in part by the QUALCOMM Research Studentship. We would especially like to thank Guy Hoffman for permitting us the use of his AUR videos. We

would also like to thank Pradipta Biswas, Ian Davies, Alison Merikangas, Stephen Murdoch, Marc Nobile, Rachel Schwartz, Graham Titmus, and Phil Tuddenham.

References

- [1] S. Baron-Cohen and S. Wheelwright. The empathy quotient: An investigation of adults with asperger syndrome or high functioning autism, and normal sex differences. *J Autism Dev Disord.*, 34:163–175, Apr. 2004.
- [2] C. Columbus. Bicentennial Man, Dec 1999. Perf. R. Williams.
- [3] M. de Weid, P. P. Goudena, and W. Matthys. Empathy in boys with disruptive behavior disorders. *J Child Psychol Psychiatry*, 46, 2005.
- [4] Duplo2100. Roomba, May 2006. Youtube. <http://www.youtube.com/watch?v=HqhIMFQNGCg>.
- [5] J. Goetz, S. Kiesler, and A. Powers. Matching robot appearance and behavior to tasks to improve human-robot cooperation. *IEEE Sym. on Robot and Human Interactive Communication (RO-MAN)*, 2003.
- [6] O. Golan, S. Baron-Cohen, J. J. Hill, and Y. Golan. The “reading the mind in films” task: Complex emotion recognition in adults with and without autism spectrum conditions. *Soc Neurosci*, 1(2), 2006.
- [7] A. I. Goldman. *Simulating Minds: The philosophy, psychology, and neuroscience of mindreading*. Oxford University Press, 2006.
- [8] D. Hannan. Coral sea dreaming, Oct. 1999.
- [9] P. J. Hinds, T. L. Roberts, and H. Jones. Whose job is it anyway? a study of human-robot interaction in a collaborative task. *Hum.-Comput. Interact.*, 19(1):151–181, 2004.
- [10] C.-C. Ho, K. F. MacDorman, and Z. A. Pramono. Human emotion and the uncanny valley: a glm, mds, and isomap analysis of robot video ratings. In *ACM Conf. on Human robot interaction*, 2008.
- [11] G. Hoffman. The robot AUR plays a character role in “the confessor”. MIT Media Lab, May 2007.
- [12] G. Hoffman. AUR robotic desk lamp - autonomous desktop assistance. MIT Media Lab, Aug. 2008.
- [13] S. Krach, F. Hegel, B. Wrede, G. Sagerer, F. Binkofski, and T. Kircher. Can machines think? int. and perspective taking with robots investigated via fmri. *PLoS ONE*, 3(7):2597, July 2008.
- [14] E. J. Lawrence, P. Shaw, D. Baker, S. Baron-Cohen, and A. S. David. Measuring empathy: reliability and validity of the empathy quotient. *Psychol Med*, 34(05):911–920, 2004.
- [15] A. Meltzoff and R. Brooks. ‘Like me’ as a building block for understanding other minds: Bodily acts, attention, and intention. In *Intention and intentionality*. MIT Press, 2001. Malle, F. and Moses, L. and Baldwin, D., eds.
- [16] S. Miller. I, Roomba, May 2007. Perfs. J. Grimm and N. Perleos.
- [17] M. Mori. Bukimi no tami. *Energy*, 7(4):33–35, 1970.
- [18] S. J. Muncer and J. Ling. Psychometric analysis of the empathy quotient (eq) scale. *Pers Individ Differ.*, 40(6):1111–1119, 2006.
- [19] S. D. Preston and F. B. M. de Waal. Empathy: Its ultimate and proximate bases. *Behav. Brain Sci.*, 25(1):1–20, 2002.
- [20] L. D. Riek and P. Robinson. Robot, rabbit, or red herring? Societal acceptance as a function of classification ease. In *IEEE RO-MAN Workshop on Robots as Social Actors*, 2008.
- [21] B. Robins, K. Dautenhahn, and R. te Boekhorst. Robots as Assistive Technology - Does Appearance Matter? In *IEEE RO-MAN*, 2004.
- [22] J. Smight. The Twilight Zone: The Lonely, Nov 13 1959. Perfs. J. Marsh and J. Warden.
- [23] L. Tirard. Molière, 2007. Perf. R. Duris.
- [24] F. Truffaut. The 400 blows, 1960. Perf. J. Leaud.
- [25] J. C. Turner. Social comparison, similarity and ingroup favouritism. In H. Tajfel, editor, *Differentiation between social groups: Studies in the social psychology of intergroup relations*. Academic Press, 1978.
- [26] B. Whitby. Sometimes it's hard to be a robot: A call for action on the ethics of abusing artificial agents. *Interact. Comput.*, 20(3), 2008.
- [27] L. Wispè. *The psychology of sympathy*. Plenum Press, 1991.