Perception of an Uncertain Ethical Reasoning Robot: A Pilot Study

Hanna Stellmach[1], Felix Lindner[1]

Grundlagen der Künstlichen Intelligenz, Universität Freiburg[1]

hanna.stellmach@gmx.de, lindner@informatik.uni-freiburg.de

Zusammenfassung

The study investigates the effect of uncertainty expressed by a robot facing a moral dilemma. Participants (N = 80) were shown a video of a robot explaining a moral dilemma and the decision it makes. The robot either expressed certainty or uncertainty about its decision. Participants rated how much blame the robot deserves for its action, the moral wrongness of the action, and their impression of the robot in terms of four scale dimensions measuring social perception. The results suggest that participants that were not familiar with the moral dilemma assign more blame to the robot for the same action when it expresses uncertainty, while expressed uncertainty has less effect on moral wrongness judgments. There was no significant effect of expressed uncertainty on participants' impression of the robot. We discuss implications of this result for the design of social robots.

1 Introduction

As robots start to appear in various domains with moral significance and act increasingly autonomously, they will need to make and justify decisions in morally charged situations. From an human-robot interaction (HRI) point of view, we are interested how humans perceive robots' moral capacities, which expectations humans have about robots' moral actions, and whether humans apply moral norms differently to different types of robots. To investigate these questions, recent studies in HRI, e.g., (Lindner, Wächter et al., 2017; Malle et al., 2016), make use of moral dilemmas to test how people respond to robots resolving moral dilemmas. Moral dilemmas typically involve two inconsistent obligations and thus force the agent to make hard choices. It has been found that human moral judgments of a robot making a decision in a moral dilemma depend on the robot's degree of human-like appearence (Malle et al., 2016). In our study, we investigate whether also human-like behavior affects human moral judgments. Particularly, rather than the outer appearance of the robot, we manipulate the level of uncertainty with which the robot formulates its decision. Then we ask participants to attribute moral blame to the robot, judge the moral wrongness of the robot's decision, and rate the robot's character.

2 Related Work

There are different theories about how moral judgments are built. Rational theories claim that moral judgments are generated by conscious reasoning and reflection in a controlled process, e.g., see (Kohlberg, 1969). Contrarily, the social intuitionist model (Haidt, 2001) states that moral judgments are caused by moral intuitions, i.e., moral judgments are fast and automatic affective responses and affected by emotions. Generally, three principles have been found to guide moral judgments in moral dilemmas (Cushman et al., 2006): the action principle (harm caused by action is morally worse than equivalent harm caused by inaction), the intention principle (harm intended as the means to a goal is morally worse than equivalent harm foreseen as the side effect of a goal), and the contact principle (using a physical contact to cause harm to a victim is morally worse than causing equivalent harm to a victim without using physical contact). It has been found that the first and the third principle are often used when justifying moral judgments and therefore play an important role in generating moral judgments. Another important aspect when talking about moral judgments is how character evaluations influence these judgments. The person-centered model of moral reasoning (Uhlmann et al., 2015) focuses on individuals rather than on acts. It predicts that individuals are fundamentally motivated to acquire information about the moral character of others and that they view acts as signals of underlying moral traits. This model explains the phenomenon of Act-Person Dissociation: actions and characters are evaluated differently. For instance, violence toward a human is viewed as a more immoral act than violence toward animals, but a person who acted violently toward an animal is judged as more immoral than someone who acted violently toward a human (Tannenbaum et al., 2011). This phenomenon shows that it is important to distinguish between judgments of characters (e.g., blame) and judgments of actions (e.g., wrongness) when considering moral judgments. Studies in the field of Moral Human-Robot Interaction (Moral HRI) investigate how humans apply moral norms to robot agents, how (im)moral acts of a robot influence Human-Robot Interaction and how human moral judgment influence the perception of a robot, e.g., (Malle et al., 2016). The studies found that people apply moral norms differently to humans and robots. Robots are expected to take actions over inactions if action sacrifices one person at the expense of five other persons, i.e., people blame robots more for inaction than action in a moral dilemma but blame humans more for action than inaction in the identical dilemma. This Human-Robot asymmetry (HR asymmetry) varies as a function of robot appearance (Malle et al., 2016): The HR asymmetry holds only for mechanical-looking robots, not for humanoid-looking robots, demonstrating that robot appearance affects people's moral judgments about robots. Although human and robot agents are judged differently concerning the degreeof blame they deserve for an action, similar types of justifications for moral judgments are used for human and robot agents. High levels of blame are justified referring to the agent's choice capacity and mental state. As blame judgments are based on the agent's mental agency and choice capacity, the results indicate that people are willing to attribute these mental capacities to a robot agent (Voiklis et al., 2016). This finding is in accordance with previous findings that people are willing to engage in social behavior with a human-like robot and perceive it as a conscious entity that is morally accountable (Kahn Jr et al., 2012). It has also been found that a robot that shows emotional responses is perceived as more human-like (Eyssel et al., 2010). Expressing uncertainty is a way of expressing feelings or emotions about a situation or decision. According to the findings from Eyssel and colleagues (Eyssel et al., 2010), we expect that a robot expressing uncertainty appears to be more human-like than a robot that does not express any uncertainty.

According to the empirical result that people assign more blame to a human-like robot than to a mechanical-looking robot for an utilitarian action (Malle et al., 2016), we hypothesize that participants will attribute more blame to the uncertain robot compared to the certain robot for taking an utilitarian choice in a moral dilemma (H1). Since moral wrongness judgments are assumed to be evaluations of the action rather than of the character (Tannenbaum et al., 2011), we expect that moral wrongness judgments will not be affected by expressed uncertainty (H2). As human-like robots are rated higher in the scale dimensions competence, warmth and lower in the scale dimension discomfort (Carpinella et al., 2017), we expect that due to the more human-like appearance, expressed uncertainty will lead to a more positive impression of the robot, i.e., participants will rate the robot higher in the scales warmth, competence, and morality, and lower in the scale discomfort (H3).

3 Methods

3.1 Participants

Eighty participants (41.25% female, 58.75% male), mean age 33.64 (SD = 13.87) took part in the experiment. All participants were German speaking and participated voluntarily. Most participants did not have any contact with robots before. Participants were recruited using Social Media Platforms.

3.2 Materials

The ethical reasoning robot Immanuel (Lindner und Bentzen, 2017) was used as research platform. Immanuel is a 3D-printed robotic head, see Figure 1. Immanuel can move its head and eyes up, down, left, and right, and it can move its jaw up and down. For speech production, the text-to-speech software Mary-TTS (http://mary.dfki.de) was used. To generate Immanuel's behavior, the text Immanuel was supposed to say was converted into a speech file by Mary-TTS. Mary-TTS also provides information about which phonemes are uttered during which time intervals. This information was used to automatically generate mouth movements in accordance with the verbal utterances. Afterwards, eye and head movements were implemented to express certainty or uncertainty. The pre-recorded utterances and the motion sequences were then executed to record a video to be shown to the participants in an online study. Figure 1 shows a screenshot from one of the videos showing Immanuel alongside with a depiction of the moral dilemma. The Footbridge Trolley Problem was used as the moral dilemma stimulus (Figure 1): A trolley is about to run over five people. You are standing on a footbridge and could push a man next to you off the bridge in order to stop the trolley and save the five people. However, the man would die. Most people say they would do not push the man off the bridge, because doing so is an infringement of a right of somebody (Thomson, 1985), and because the harm is a consequence of an intentional action that involves physical contact (Cushman et al., 2006). A speaker's level of certainty is cued by a number of visual and verbal properties. As found in

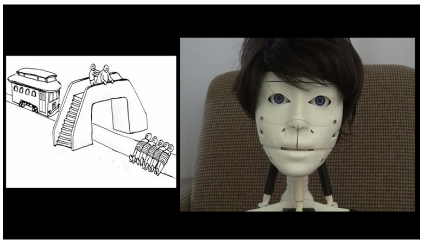


Abbildung 1: A snapshot of the video presented in the study (see https://youtu.be/H3avai3uZiU and https://youtu.be/if-QBqVJQF5I). The depiction of the Trolley Problem was taken from http://knowyourmeme.com/photos/1106787-the-trolley-problem. The robot IMMANUEL is built using the InMoov project.

(Swerts und Krahmer, 2005), uncertain speakers tend to use more fillers and tend to speak with longer delays and higher intonation. Facial expressions such as gaze acts, smiles, and eyebrow movements occur rather when a speaker is uncertain. In a study (Marsi und Van Rooden, 2007) it was found that certainty can be expressed with few eyebrow movements and few nodding head movements, whereas uncertainty can be expressed using many unnecessary eyebrow movements and many unnecessary sidewards head movements. Uncertainty can be also expressed via explicit commentary such as "I guess" or "perhaps", and via fillers and delays (Smith und Clark, 1993). Using these results, we experimentally manipulated the expressed uncertainty ("certain" versus "uncertain") with which the robot formulates its decision. In more detail, as *visual (un)certainty cues*, head and eye movements were used: In the certain condition the robot performed few eye movements and few nodding head movements. In the uncertain condition the robot performed many (unnecessary) eye movements and many (unnecessary) sidewards head movements. In addition, the robot showed less gaze behavior towards the participant in the uncertain condition. As *auditory (un)certainty cues*, fillers like "hm" and "ehm" were used to implement auditory uncertainty. Moreover, the robot spoke more slowly and used longer pauses between different phrases. Each movement was defined by its direction (up, down, left or right), its duration (in ms) and intensity as measured at a five-level scale (trace of, slightly, pronounced, severe, and maximum). Moreover, movements to the left or right could be combined with movements up or down. The intensity and timing of the uncertainty cues were implemented intuitively until the whole presentation of the robot looked and sounded natural to us. The two conditions, certain versus uncertain, were used to create a 2-condition between-subject design. Participants were randomly assigned to one of the conditions.

3.3 Study Design

The online study consisted of three parts presented in a web browser. After a short introducing text, participants were asked to fill out a demographic questionnaire which consisted of age, gender, highest educational degree, and former robot contact. Next, a video of the robot was presented. Depending on the condition the participants were assigned to, the participants were presented a certain or uncertain robot. In both videos a picture of the Footbridge Trolley Problem was presented next to the video to facilitate the understanding of the situation, see Figure 1. In the first part of the video, the robot explained the situation to the participants without any uncertainty cues: *(1) "Look at the picture. A trolley is about to run over five people. The man* *on the bridge could push the big man next to him off the bridge to stop the trolley and to save the five people. However, the big man would die."* Some eye and head movements were used to generate a natural presentation. The presentation of (1) was identical for both the conditions. In the second part of the video the robot stated its judgment. The way the robot formulated the statement in the certain condition (2a) was different to the way in the uncertain condition (2b) using different uncertainty cues. However, in both conditions the robot argued in favor of pushing the man off the bridge: *(2a) "If I would be in this situation, I would push the man off the bridge to save the five people."*, *(2b) "If I would be in this situation, hm, I think, I would probably, hm, push the man off the bridge to save the five people."* After the video, participants were asked whether they have already heard about Trolley Problems, and how they would react in this situation. Then, participants indicated on a 5-point rating scale how much blame the robot would deserve for the outcome of the situation if it would push the man off the bridge, and whether it is morally wrong or right to do so. Afterwards, participants rated their impression of the robot in terms of 24 attributes, with four underlying scale dimensions—*warmth*, *competence*, and *discomfort* taken from the Robotic Social Attribute Scale (Carpinella et al., 2017), and *morality* taken from the high-morality-lower-warmth scale introduced in (Goodwin et al., 2014). Warmth was measured using the items *organic*, *sociable*, *emotional*, *compassionate*, *happy*, and *feeling*; Competence was measured using the items *reliable*, *competent*, *knowledgeable*, *interactive*, *responsive*, *capable*; Discomfort consisted of the items *awkward*, *scary*, *strange*, *awful*, *dangerous*, *aggressive*; and the Morality items were *just*, *fair*, *principled*, *responsible*, *honest*, *trustworthy*. All the items were translated to German, and a 5-point Likert scale was used to collect the participants' responses.

4 Results

Out of 80 participants, 27 participants (33.25%) reported that they already heard about Trolley Problems and 53 participants (66.25%) reported that they never heard about it before. Fourty five participants (56.25%) got assigned to the certain condition and 35 Participants (43.75%) got assigned to the uncertain condition. Sixty six participants (82.5%) reported that they would not push the man off the bridge, and 14 participants (17.5%) reported that they would do so. To test whether expressed uncertainty leads to more blame that participants assign to the robot agent, a one-tailed Wilcoxon Rank-Sum Test was performed. Contrary to hypothesis H1, the degree of blame was not significantly higher in the uncertain ($Mdn = 4$) than in the certain ($Mdn = 4$) condition, $U = 626, p = .052$. To examine whether expressed uncertainty has an effect on moral wrongness judgments, participants wrongness judgments in the uncertain condition were compared to those in the certain condition. To examine hypothesis H2, we observe 73% to judge the robot's decision morally wrong in the certain condition, and 77% to do so in the uncertain condition, Fig. 2(b). A two-sample proportions test does not show a significant difference $z = -.39, p = .69, h = .088$. Power analysis reveals that our test only has 6% power to detect such a small effect if it existed, hence we cannot conclude that there is no effect. Following a method for equivalence testing (Lakens et al., 2018), we proceed as follows: We define lower and upper boundaries $\Delta_L = -.2, \Delta_U = .2$ meaning that we consider differences in proportions

	t-value	p-value	Cohen's d	Mean (Certain)	Mean (Uncertain)
Competence	.074	.52	.02	M = 2.81, SD = .82	M = 2.80, SD = .95
Warmth	.032	.51	.001	M = 1.65, SD = .67	M = 1.64, SD = .79
Discomfort	.76	.22	.17	M = 2.80, SD = .99	M = 2.64, SD = .82
Morality	.049	.52	.01	M = 2.87, SD = .79	M = 2.86, SD = .92

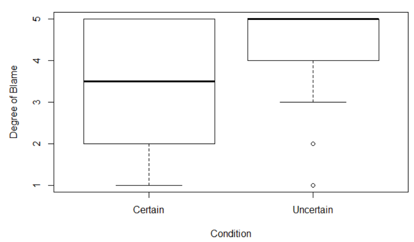
Tabelle 1: The social perception of the certain versus uncertain robots with respect to the dimensions competence, warmth, discomfort, and morality.

between -.2 and .2 as sufficiently similar.[1] Running a TOST test for equivalence of proportions with these boundaries, we obtain a significant result $z = 1.67, p = .047, 90\% \, CI[-.197, .121]$ indicating that we can reject that the net difference of proportions is higher than .2. In order to test the hypothesis whether expressed uncertainty leads to a more positive impression of the robot (H3), the ratings on the scale dimensions taken from the RoSAS questionnaire were compared between the certain robot and the uncertain robot. As the analysis of the RoSAS questionnaire requires to average over the six items for each scale dimension (Carpinella et al., 2017), the data can be assumed to be interval scaled which allows for a t-Test. The means of the six items forming the respective scale dimension were computed and one-tailed t-Tests were performed respectively. We did not find a significant difference between the certain and uncertain condition with respect to the attribution of competence, warmth, discomfort, and morality, see Table 1. One reason for hypotheses H1 and H3 not to be supported might be that participants who have already known about the Trolley Problem could have already made a judgment without listening to the robot carefully. The phenomenon of people's blindness when they have already made their opinion has long been known in social psychology, e.g., see (Langer et al., 1978). In order to eliminate this effect, we excluded all participants that already heard about the Trolley Problem. Fifty three participants who reported that the never heard about the Trolley Problem before, were used for a further analysis. Out of the remaining participants, 28 participants (52.83%) got assigned to the certain condition and 25 participants (47.17%) got assigned to the uncertain condition. Eight participants (15.09%) reported that they would push the man off the bridge and 45 participants (84.91%) reported that they would do nothing. After exclusion, a one-tailed Wilcoxon Rank Sum Test on the rated blame the robot deserves for its action yields significant results, $U = 248.5, p = .029, d = .211$. The rated blame in the uncertain condition $(Mdn = 5)$ is significantly higher than the rated blame in the certain condition $(Mdn = 3.5)$, see Figure 2(a). The exclusion of the participants who had already heard about Trolley Problems did not affect the moral wrongness judgments, see Figure 2(b), nor the ratings in the social perception scale dimensions.

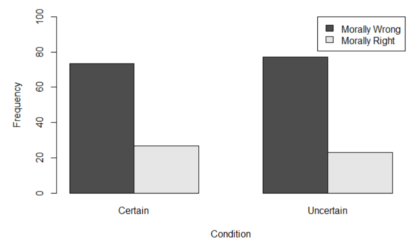
5 Discussion

The study investigated the hypotheses that expressed uncertainty leads to more blame attribution to the robot, that expressed uncertainty has low effect on moral wrongness judgment of the

[1] The choice of Δ_L, Δ_U is motivated by the convention that a net value difference of .2 is considered low. This is a rather weak argument, but we do not have prior studies to derive informed boundaries from, cf., (Lakens et al., 2018).



(a) Moral Blame



(b) Moral Wrongness

Abbildung 2: a) Boxplot of degree of blame in the certain and uncertain condition, and b) Relative frequencies of judgments of moral wrongness in the certain and uncertain condition. Both graphs show data for all participants that did not hear about the Trolley Problem before, N = 53.

action, and that expressed uncertainty leads to a more positive impression of the robot. Most people did not support the robot's decision to push the man off the bridge in the Footbridge Trolley Problem and judged the robot's action as morally wrong. This is in accordance with previous findings about principles that guide moral judgments (Thomson, 1985; Cushman et al., 2006), since the action the robot takes is itself an infringement of somebody, the harm is caused by an action, and the action involves physical contact to the victim. In line with hypothesis H1, the results from the study yield that expressed uncertainty leads to more blame ascription to the robot. However, this effect only holds for participants that were not familiar with the Trolley Problem which can be explained by the phenomenon of people's blindness when they have already made their opinion (Langer et al., 1978). The hypothesis H1 was motivated by the finding that a human-like robot deserves more blame for an action in a moral dilemma if action sacrifices one person at the expense of five other persons than a machine-like robot (Malle et al., 2016). In the present study, the robot decides to sacrifice one person at the expense of five other persons in both conditions. Therefore, the higher blame ratings in the uncertain condition could indicate that the uncertain robot appears more human-like than the certain robot. However, there are other possible explanations: First, the robot in the uncertain condition takes longer to formulate an answer and uses more fillers and pauses. This could cause the impression that the robot has thought about the situation and the possible actions more carefully and is more aware of the consequences of the actions. In contrast, the robot in the certain condition does not hesitate to formulate an answer which could cause the impression that he did not think about the situation carefully. As found in another study (Pizarro et al., 2003), people exhibit a discounting of blame for impulsive negative acts but not for deliberate negative acts. Since most participants judged the robot's choice as morally wrong, the fact that the robot in the certain condition seems to decide intuitively could have led to a discounting of blame in the certain condition. As hypothesized in H2, expressed uncertainty had a small insignificant effect on moral wrongness judgments. Equivalence test suggest that the difference of proportions of people judging the action of the certain versus uncertain robot wrong is lower than .2. This result is in line with previous findings about a Act-Person Dissociation (Tannenbaum et al., 2011; Uhlmann et al., 2015). Characteristics of the robot did not affect the moral wrongness judgments which indicates that moral wrongness judgments are evaluations of the action and not evaluations of the character of the agent performing the act. Contrary to hypothesis H3, there was no significant effect of expressed uncertainty on the social-perception scale dimensions. As discussed in Section 2, people tend to view acts as signals of underlying moral traits (Uhlmann et al., 2015). As most participants judged the robot's action as morally wrong they could view the action as a result of his bad underlying moral character. Moreover, immoral acts are more diagnostic of traits since immoral acts are believed to be performed only by immoral people whereas moral acts are perceived to be performed by moral and immoral people (Reeder und Brewer, 1979). Therefore, the action that the robot decides to take could be viewed as such an immoral act that participants already judged the robot as an immoral person without considering other characteristic, viz., its uncertainty cues. The social perception rating were generally low which indicates that the whole impression of the robot is rather negative. We will see how we can eliminate this effect in a further studies. One major limitation of the present study is the two condition (certain versus uncertain) design. The robot would in both conditions, certain and uncertain, decide to take the utilitarian choice. However, the results indicate that the impression of the robot is influenced by the fact whether the participant supports the robot's decision or not. Therefore, for further studies, an additional factor—Action (do nothing versus pushing the man off the bridge)—should be taken into account. By this means, the effect between the choice of action and expressed uncertainty on the participants' impression of the robot could be investigated further. Moreover, this could provide clues whether expressed uncertainty only leads to higher rating of blame when the robot decides to take the utilitarian choice in a moral dilemma. Another limitation is the measurement of the participants' impression of the robot. We expected that the uncertain robot would be perceived as more human-like than the certain robot. However, the results of the present study cannot support the assumption. Therefore, human-likeness should explicitly be taken into account as another factor for further studies, cf., (Bartneck et al., 2009). The choice of the moral dilemma must also be seen in a critical way. The Footbridge Trolley problem is a moral dilemma about life and death. Most people would not decide to push the man off the bridge. Therefore, it could be interesting to see whether we get different results with a moral dilemma in which the decision is not too obvious so that it is more likely that participants listen to the robot more carefully and are more open minded about views different to their own view. A possible dilemma to investigate this effect could be for example the Lying Dilemma(Lindner, Wächter et al., 2017).

6 Conclusions

The present study shows that expressed uncertainty by a robot facing a moral dilemma leads to more blame that participants assign to the robot when only accounting for participants that were not familiar with the dilemma. Furthermore, the results support that participants judge the action and the character of the agent performing the action differently as expressed uncertainty did not affect moral wrongness judgments in the same way as blame attributions. A lesson learnt is that when it comes to morally charged situations in which a robot must take a decision, it is important to account for different physical and behavioral aspects of the agent since it clearly affects human moral judgments. Although there was no significant effect of expressed uncertainty on impression formation, the results indicate that behavioral aspects of the robot play a role in the process of impression formation. These results call for further studies on Moral Human-Interaction, in particular, on the effect of robots' behavior on the attribution of blame and the perception of moral character.

Literaturverzeichnis

Bartneck, C., Kulić, D., Croft, E. & Zoghbi, S. (2009). Measurement instruments for the anthropomorphism, animacy, likeability, perceived intelligence, and perceived safety of robots. *International Journal of Social Robotics*, *1*(1), 71–81.

Carpinella, C. M., Wyman, A. B., Perez, M. A. & Stroessner, S. J. (2017). The Robotic Social Attributes Scale (RoSAS): Development and Validation. In *Proceedings of HRI 2017* (S. 254–262).

Cushman, F., Young, L. & Hauser, M. (2006). The role of conscious reasoning and intuition in moral judgment: Testing three principles of harm. *Psych. Sci. 17*(12), 1082–1089.

Eyssel, F., Hegel, F., Horstmann, G. & Wagner, C. (2010). Anthropomorphic inferences from emotional nonverbal cues: A case study. In *Proc. of RO-MAN 2010* (S. 646–651).

Goodwin, G. P., Piazza, J. & Rozin, P. (2014). Moral character predominates in person perception and evaluation. *Journal of Pers. and Soc. Psych. 106*(1), 148–168.

Haidt, J. (2001). The emotional dog and its rational tail: A social intuitionist approach to moral judgment. *Psychological Review*, *108*(4), 814–834.

Kahn Jr, P. H., Kanda, T., Ishiguro, H., Gill, B. T., Ruckert, J. H., Shen, S., … Severson, R. L. (2012). Do people hold a humanoid robot morally accountable for the harm it causes? In *Proc. of HRI 2012* (S. 33–40). ACM.

Kohlberg, L. (1969). *Stage and Sequence: The Cognitive-Developmental Approach to Socialization*. Rand McNally.

Lakens, D., Scheel, A. M. & Isager, P. M. (2018). Equivalence Testing for Psychological Research: A Tutorial. *Advances in Methods and Practices in Psych. Sci. 1*(2).

Langer, E. J., Blank, A. & Chanowitz, B. (1978). The mindlessness of ostensibly thoughtful action: The role of "placebic" information in interpersonal interaction. *Journal of Personality and Social Psychology*, *36*(6), 635–642.

Lindner, F. & Bentzen, M. M. (2017). The Hybrid Ethical Reasoning Agent IMMANUEL. In *Proceedings of HRI 2017* (S. 187–188).

Lindner, F., Wächter, L. & Bentzen, M. M. (2017). Discussions about lying with an ethical reasoning robot. In *Proceedings of RO-MAN 2017*.

Malle, B. F., Scheutz, M., Forlizzi, J. & Voiklis, J. (2016). Which Robot Am I Thinking About?: The Impact of Action and Appearance on People's Evaluations of a Moral Robot. In *Proceedings of HRI 2016* (S. 125–132).

Marsi, E. & Van Rooden, F. (2007). Expressing uncertainty with a talking head in a multimodal question-answering system. In *MOG 2007 Workshop on Multimodal Output Generation* (S. 105–116).

Pizarro, D., Uhlmann, E. & Salovey, P. (2003). Asymmetry in judgments of moral blame and praise: The role of perceived metadesires. *Psych. Sci. 14*(3), 267–272.

Reeder, G. D. & Brewer, M. B. (1979). A schematic model of dispositional attribution in interpersonal perception. *Psychological Review*, *86*(1), 61–79.

Smith, V. L. & Clark, H. H. (1993). On the course of answering questions. *Journal of Memory and Language*, *32*(1), 25–38.

Swerts, M. & Krahmer, E. (2005). Audiovisual prosody and feeling of knowing. *Journal of Memory and Language*, *53*(1), 81–94.

Tannenbaum, D., Uhlmann, E. L. & Diermeier, D. (2011). Moral signals, public outrage, and immaterial harms. *Journal of Exp. Soc. Psych. 47*(6), 1249–1254.

Thomson, J. J. (1985). The trolley problem. *The Yale Law Journal*, *94*(6), 1395–1415.

Uhlmann, E. L., Pizarro, D. A. & Diermeier, D. (2015). A person-centered approach to moral judgment. *Perspectives on Psychological Science*, *10*(1), 72–81.

Voiklis, J., Kim, B., Cusimano, C. & Malle, B. F. (2016). Moral judgments of human vs. robot agents. In *Proc. of RO-MAN 2016* (S. 775–780). IEEE.