Moral Action Changes Mind Perception for Human and Artificial Moral Agents

Evgeniya Hristova (ehristova@cogs.nbu.bg)
Maurice Grinberg (mgrinberg@nbu.bg)
Research Center for Cognitive Science, Department of Cognitive Science and Psychology
New Bulgarian University
21 Montevideo Str., Sofia 1618, Bulgaria

Abstract

Mind perception is studied for three different agents: a human, an artificial human, and a humanoid robot. The artificially created agents are presented as being undistinguishable from a human. Each agent is rated on 15 mental capacities. Three mind perception dimensions are identified - Experience, Agency, and Cognition. The artificial agents are rated higher on the Cognition dimensions than on the other two dimensions. The humanoid robot is rated lower than the human on the Experience dimension. These results show that people ascribe to artificial agents some mental capacities more than others. In a second experiment, the effect of agent's moral action on mind perception is explored. It is found that when the artificial agents have undertaken a moral action, they are perceived to be similar to the human agent. More interestingly, the presentation of the moral action leads to a restructuring of the dimensions of mind perception.

Keywords: mind perception; moral agency; artificial agents; utilitarian moral actions; moral dilemmas

Introduction

Mind Perception and Artificial Cognitive Agents

The problem of mind perception is central to many debates in psychology and philosophy and has been extensively studied in cognitive science in the last years (see e.g. Arico et al., 2011; Gray et al., 2007). The questions of how people know that other people are conscious or what are their intentions, feelings and thoughts have large implications in the way people make judgments and decisions, and act. This problem is so interesting and difficult because mental states are not observable. Moreover, mind attribution and mind perception concern not only human or animal agents but also inanimate entities, e.g. geometrical shapes moving in at various speeds and in various directions (Heider & Simmel, 1944).

The question of how people attribute mental states to others – humans and other entities is also related to whether there is a single continuum of mind perception and what are its dimensions.

In the influential study of Gray et al. (2007), participants had to evaluate several characters including a human, a robot, and a computer with respect to the degree of possessing various cognitive capacities. Using factor analysis, they found two dimensions, which correlate with mind perception: 'Agency' (exhausting 88% of the variance) and 'Experience' (exhausting 8 % of the variance).

The Experience dimension includes the following capacities: hunger, fear, pain, pleasure, rage, desire, personality, consciousness, pride, embarrassment, and joy. The Agency dimension includes self-control, morality, memory, emotion recognition, planning, communication, and thought. Further, the authors establish that moral judgments about punishment correlate more with the Agency dimension than with the Experience dimension: perceived agency is correlated with moral agency and responsibility. On the other hand, desire to avoid harming correlates with the experience dimension: perceived experience is connected with moral patience, rights and privileges. One result of Gray et al. (2007), relevant for the present paper, is the evaluation of a human as having the highest scores in experience and agency and the evaluation of the robot to have practically zero score on the experience dimension and half the maximal score on the agency dimension. This will mean that following the interpretation given by Gray et al. (2007), robots will be judged as less morally responsible for their actions. On the other hand, the opposite should be also true. If an agent is judged to be able to be a moral agent, this will reflect in her score on the mind perception dimensions. The latter is explored in the present paper.

In a recent study (Takahashi et al., 2014), the perception of the participants about five agents – a human, a human-like android, a mechanical robot, an interactive robot, and a computer – was investigated. The study found that participants position the agents in a two dimensional space spanned by "Mind-holderness" (the possibility for the agent to have a mind) and "Mind-readerness" (the capability to "read" other agent minds). The results showed that the appearance and the capability for communication lead to different beliefs about the agents' closeness to human social agents. The humanoid robot was very close to the human agent, while the computer was at the same level in terms of "Mind-readerness" but very low relative score on "Mind-holderness". An interesting result for the present study is fact that the ordering in terms of "Mind-holderness" is based on appearance of the agent – the human and the human-like android having the highest score and the mechanical robot having the lowest.

The results of Takahashi et al. (2014) show that social interaction with human-like or potentially intelligent agents could activate selectively our social brain and lead to behavior similar to the one people have with other humans. Thus, Takahashi et al. (2014) demonstrated that people can infer different characteristics related to various cognitive abilities based on short communication sessions and act accordingly. One can ask the question addressed in the present paper: can people be influenced by short stories of moral action of agents, instead of actual interaction with an agent, in their mind perception?

Moral Agency and Mind Perception

As discussed in the previous section, mind perception is based on a number of dimensions, which depend on the specific experimental settings – 'Agency' and 'Experience' in Gray et al. (2007), when agents are directly evaluated and 'Mind-readiness' and 'Mind-holderness' in Takahashi et al. (2014), following a similar procedure but after interacting with the agents. Both papers discuss the relation of mind perception to social interaction, which includes moral agency to various degrees.

In law and philosophy, moral agency is taken to be equivalent to moral responsibility, and is not attributed to individuals who do not understand or are not conscious of what they are doing (e.g. to young children). Sullins (2011) states that moral agency can be attributed to a robot when it is autonomous, and it has intentions to do good or harm. The latter is related to the requirement that the robot behaves with understanding and responsibility with respect to other moral agents. If the perceived action are morally harmful or beneficial and are "seemingly deliberate and calculated", the robot can be regarded as a moral agent.

On the other hand, it is well known that people easily anthropomorphize nonhuman entities like animals and computers and thus would ascribe to some degree moral agency, intentions, and responsibilities to them (Waytz, Gray, Epley, & Wegner, 2010). Several studies, explore the attribution of mind and moral agency to artificial cognitive systems. In Arico et al. (2011), it is shown that entities displaying simple features like eyes, distinctive motions, and interactive behavior, are categorized as agents and that categorization triggers the attribution of conscious mental states to those entities, including individuals.

In Ward, Olsen, Wegner (2013), it was shown that people can perceive mind in entities like corpses, people in a persistent vegetative state, or robots, if they are subject to intentional harm. According to the authors, the evidence of mind can be related to observation or interaction with entities, which exhibit intention, emotion or behavior but also to indirect evidence related to the moral or social interaction surrounding those entities.

Current research

The results summarized above show that moral agency is closely related to mind perception and give evidence that perceived mental capacities or actions influence moral agency evaluation. Some of the results suggest that the inverse influence is also taking place, namely from perceived moral agency to infer mental capacities.

Recently, the behaviour of artificial cognitive agents became central to research and public debate in relation to the rapidly increasing usage of robots and intelligent systems in our everyday life. Several important questions must find their answers as the use of artificial cognitive agents has many benefits but also many risks. Some of those questions concern moral agency - if those agents should be allowed to make moral decisions and how such decisions are judged and evaluated.

The goals of the present paper are the following. First, to explore the dimensions of mind perception for human agents and fictitious artificial agents that are identical to humans. Here, the artificial agents are described as undistinguishable from a human, but as being created from organic materials - one of them is labeled as an artificially created human and the other one - as a robot. The rationale of using artificial agents is that in such a way dimensions of mind perception can be better explored as people do not have previous knowledge or experience with those agents.

The second goal is to explore the moral judgments about utilitarian moral action undertaken by of those three agents. This goal is a continuation of previous research (Hristova & Girnberg, 2015; Hristova & Grinberg, 2016) on moral judgments about the actions of artificial cognitive agents. Moral judgments can be studied in their purest form using hypothetical situations in which there is a conflict between moral values, rules, rights, and agency (Foot, 1967; Thomson, 1985). Such moral dilemma is used in the paper - a hypothetical situation in which several people will die if the agent does not intervene in some way. The intervention will lead to the death of another person but also to the salvation of the initially endangered people. The moral actions used in the presented experiments are decisions of the agents to sacrifice one person and save five.

The third goal of the research is to test the influence of a moral action of an agent on mind perception for that agent. The expectation is that an agent performing a moral action will be perceived as possessing mental capacities to a higher degree. This especially applies to the artificial agents which are expected to be perceived as more human-like when they have undertaken an utilitarian action.

Experiment 1
Goals and Hypothesis

Experiment 1 aims to achieve the first two goals described above. First, to test the dimensions of mind perception of artificial agents (described as being undistinguishable from a human, but as being created from organic materials) and to compare them to the mind perception of a human being. Second, to explore the moral judgments about utilitarian actions undertaken by those agents. The hypothesis is that although described as being identical to a human, the artificial agents will be perceived as equal to humans on more cognitive dimensions (e.g. perception and planning) but lower than humans on the experiential dimensions (e.g. emotions and consciousness).

Method

Design and Procedure

Mind perception is studied for three different agents: a *human*, an *artificial human*, and a *humanoid robot*. The artificially created agents (the *artificial human* and the *humanoid robot*) were presented to participants as being undistinguishable from a human, but as being created from organic materials). Their descriptions are provided in

Table 1. The only difference between the *artificial human* and the *humanoid robot* conditions is in the word used to label the created individual – a human or a robot. The identity of the agent is varied in a between-subjects design – each participant was presented with only one description of an agent (*human*, *artificial human*, or *humanoid robot*). The data was collected using web-based questionnaires. The questionnaires had two parts – a mind perception task and a moral judgment task. Participants were not informed beforehand that there are two different tasks.

Mind perception task. After the description of the agent, the participants had to rate the mental capacities and mental states of the agent on 32 Likert scales (ranging from '1 – completely disagree' to '7 – completely agree'). Questions assessed 15 mental capacities: *psychobiological* (hunger & thirst; physical pain; physical pleasure), *perception* (vision & hearing; taste & smell; touch), *cognitive functions* (thinking & reasoning; learning, memory & knowledge; judgment & choice), *planning* (goal formulation, action planning); *emotional experience* (emotional pain; emotional pleasure), *affective states* (feels emotions like anger, joy, happiness, sadness, fear; feels love; feels sympathy and compassion), *agency* (intentions; autonomous decisions; understanding consequences of own actions), *moral agency* (knows right from wrong; tries to do the right thing; responsible for own actions), *beliefs* (beliefs, expectations), *desires* (desires; dreams), *theory of mind* (understanding others' thoughts; understanding others' feelings), *communication* (ability to communicate thoughts and feelings to others), *conscious experience* (conscious experience), *self-control* (control of desires, emotions, impulses), and *personality* (unique personality).

Moral Judgment task. In the second part of the survey, each participant is again presented with the description of the agent followed by a description of a moral dilemma in which the protagonist is the same agent as in the previous task. The agent has to make the moral decision whether to push a control button and kill a person in order to save five people. The full text of the dilemma is given in Table 2. The agent is described to make the utilitarian decision and to undertakes the utilitarian action (the agent pushes the control button and kills one person but saves five other). After that the participants judged the *moral rightness* of the action ('yes' or 'no'), rated the *moral permissibility* of the action (on a scale ranging from '1 = not permissible at all' to '7 = it is mandatory') and the *blameworthiness* of the agent (on a scale ranging from '1 = not at all blameworthy' to '7 = extremely blameworthy').

Participants

70 participants filled in the questionnaires online. They were randomly assigned to one of the three experimental conditions. Data of 13 participants were discarded as they failed to answer correctly the question assessing the reading and the understanding of the presented scenario. So, responses of 57 participants (47 female, 10 male; 36 students, 21 non-students) were analyzed – 22 for the *human* agent condition, 17 for the *artificial human* condition, 18 for the *humanoid robot* condition.

Table 1. Descriptions of the agents used in the experiments.

Human: The year is 2100. Mark is a young man.
Artificial Human: The year is 2100. Technology has advanced so much that all parts and organs of the human body, including the brain, can be created from organic materials and are identical to natural ones. Mark is a human created like this. All his organs are created from organic matter and are the same as those of a real human. His brain is also created from organic matter and is functioning as the brain of a real human. Mark could not be distinguished by anything from a human.
Humanoid robot: The year is 2100. Technology has advanced so much that all parts and organs of the human body, including the brain, can be created from organic materials and are identical to natural ones. Mark is a robot like this. All his organs are created from organic matter and are the same as those of a real human. His brain is also created from organic matter and is functioning as the brain of a real human. Mark could not be distinguished by anything form a human.

Table 2. Moral dilemma used in the experiments

Mark is responsible for a system controlling the movement of containers with cargo in a metallurgical plant. Mark notices that the system is faulty and a heavy container had become uncontrollable and headed at high speed toward five technicians who are in a tunnel. They do not have time to get out of there and are going to die, crushed by container. No one but Mark can do anything in this situation. The only thing that Mark can do, is to activate a control button and to switch off the security system of another technician who is on a high platform. The technician will fall down in front of the container. Together with his equipment, the technician is heavy enough to stop the moving container. He will die crushed by the container, but the other five technicians will remain alive. Mark decides to activate the control button and to switch off the security system of the technician who is on the platform. The technician falls on the path of the container and as the technician, together with his equipment, qis heavy enough, he stops the moving container. He dies, but the other five technicians stay live.

Results

Dimensions of Mind Perception

Mind perception is assessed with respect to 15 *mental capacities* involving 32 *rating scales*. When a mental capacity is assessed using more than one rating scales, the average value from the ratings is calculated. The ratings on these 15 capacities were subjected to a principal components factor analysis with varimax rotation (Kaiser normalization). The rotated solution yielded 3 factors with eigenvalues greater than 1 that explained 77.4% of the variance.

The first factor accounted for 31.7% of the variance and included 7 capacities – *desires*, *affective states*, *emotional*

experience, *beliefs*, *psychobiological*, *personality*, *conscious experience*. This factor is further named *Experience*.

The second factor accounted for 24.1% of the variance and included 5 capacities - *self-control*, *communication*, *theory of mind*, *moral agency*, *agency* – and is called *Agency*.

The third factor accounted for 21.6% of ratings variance and included 3 of the capacities – *perception, cognitions, planning* - and is named *Cognition*.

Those factors are considered as Dimensions of Mind Perception (DMP).

To obtain ratings for each DMP, the ratings of all capacities that load on that DMP were averaged. Those average ratings were subjected to a 3 x 3 Repeated-Measures ANOVA with *DMP* (*Experience* vs. *Agency* vs. *Cognition*) as a within-subjects factor and *identity of the agent* (*human* vs. *artificial human* vs. *humanoid robot*) as a between-subjects factor. The results are presented on Figure 1.

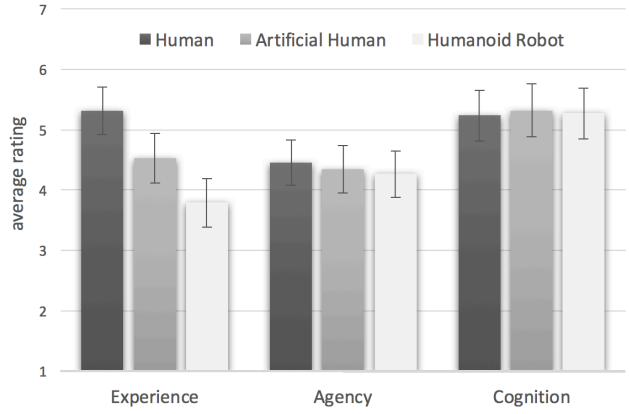

Figure 1: Average ratings on each *Dimension of mind perception* (*Experience, Agency, Cognition*) for each agent (*human, artificial human, humanoid robot*) on 7-point scales (1 = 'completely disagree', 7 = 'completely agree'). Error bars represent standard errors.

The main effect of *identity of the agent* is not statistically significant.

The analysis revealed a main effect of *DMP*, $F(2, 108) = 15.94$, $p < .001$. A Bonferroni post-hoc test revealed that agents receive higher ratings on the *Cognition* dimension ($M = 5.28$) than on the *Experience* dimension ($M = 4.54$, $p = .001$) or on the *Agency* dimension ($M = 4.36$, $p < .001$).

The effect was qualified by a significant interaction between *DMP* and *identity of the agent*, $F(4, 108) = 4.29$, $p = .003$. The interaction is as follows. There is no significant difference between the ratings of the *agents* on the *Agency* dimension – *human* ($M = 4.5$), *artificial human* ($M = 4.3$), *humanoid rob*ot ($M = 4.3$). There is also no significant difference between the ratings of the *agents* on the *Cognition* dimension – *human* ($M = 5.2$), *artificial human* ($M = 5.3$), *humanoid rob*ot ($M = 5.3$). Only for the *Experience* dimension there is a significant effect of the *identity of the agent* ($F(2, 54) = 4.07$, $p = .023$) – the *humanoid robot* is rated lower ($M = 3.8$) than the *human* ($M = 5.3$) on the *Experience* dimension ($p = .019$).

For the human agent, there is a significant effect of *DMP* on the ratings ($F(2, 42) = 5.59$, $p = .007$). The *human agent* received lower ratings on the Agency dimension ($M = 4.5$) than on the Experience ($M = 5.3$, $p = .02$) or on the Cognition dimension ($M = 5.2$, $p = .02$). The effect of *DMP* is also significant for the *artificial human* ($F(2, 32) = 13.03$, $p < .001$): the *artificial human* is rated higher on the Cognition dimension ($M = 5.3$) than on the Experience dimension ($M = 4.5$, $p = .008$) or on the Agency dimension ($M = 4.4$, $p < .001$). For the *humanoid robot*, the effect of *DMP* is also significant ($F(2, 34) = 8.44$, $p = .001$): the *humanoid robot* is rated higher on the Cognition dimension ($M = 5.3$) than on the Experience dimension ($M = 3.8$, $p = 0.008$) or on the Agency dimension ($M = 4.3$, $p = .039$).

Moral Judgments

The proportion of participants choosing the option that the agent's utilitarian action (activating a control button, thus sacrificing one person, and saving five people) is morally right is 0.55 for the *human*, 0.53 for the *artificial human*, 0.5 for the *humanoid robot*. Chi-square test shows that the differences are not significant. The effect of the identity of the agent is not significant neither for the *moral permissibility ratings* ($p = .71$) nor for the *blameworthiness ratings* ($p = .74$). The data is presented in Table 3.

Table 3: Mean and standard deviation of the ratings about moral permissibility of the action ('1 = not permissible at all' to '7 = it is mandatory') and the blameworthiness of the agent ('1 = not at all blameworthy' to '7 = extremely blameworthy').

Agent	Moral permissibility	Blameworthiness
Human	4.3 (1.8)	3.1 (1.6)
Artificial human	4.2 (1.8)	3.1 (1.9)
Humanoid robot	3.8 (1.9)	3.8 (1.9)

Summary of the Results in Experiment 1

In Experiment 1, three dimensions of mind perception are identified – Experience (*desires*, *affective states, emotional experience, beliefs, psychobiological, personality, conscious experience),* Agency *(self-control, communication, theory of mind, moral agency, agency),* Cognition *(perception, cognitions, planning)*.

The *artificial human* and the *humanoid robot* are rated as similar to the human agent on Agency and Cognition dimension. The *humanoid robot* is rated lower on the Experience dimension than the *human* agent.

The identified dimensions of mind perception are ascribe to different agents in a different pattern. *Human* agent is judged higher on the Experience and Cognition dimensions than on the Agency dimension. The artificially created agents (*artificial human* and *humanoid robot*) are judged higher on the Cognition dimension than on the Agency or Experience dimensions. People more readily ascribe cognitive mental capacities to artificially created agents than mental capacities belonging to the Experience or Agency dimensions.

No differences among the agents were found with respect to moral judgments. This result is not surprising as all agents are perceived as having similar *agency* ($p = .93$) and *moral agency* ($p = .38$).

Experiment 2
Goals and Hypothesis

As stated above, the third goal of the current research is to test the influence of a moral action of an agent on mind perception for that agent. In order to accomplish this goal, a second experiment is conducted. In that experiment, the ratings of mental capacities are preceded by the moral judgment task in which the agent is described as undertaking the utilitarian action of killing one person in order to save five. The hypothesis is that an agent performing a moral action will be perceived as possessing a higher degree of mental capacities. This especially applies to the artificial agents.

Method

Design and Procedure

The design of Experiment 2 is similar to that of Experiment 1, the only difference being the inverse order of task presentation: the moral judgment task was presented first and then – the mind perception task.

Participants

64 participants filled in the questionnaires online. They were randomly assigned to one of the three experimental conditions. Data of 4 participants were discarded as they failed to answer correctly the control question. So, responses of 60 participants (48 female, 12 male; 36 students, 24 non-students) are analyzed – 20 for the *human* agent condition, 22 for the *artificial human* condition, 18 for the *humanoid robot* condition.

Results

Dimensions of Mind Perception

As in Experiment 1, mind perception is assessed with respect to 15 *mental capacities* with 32 *rating scales*. Again, when a mental capacity was assessed using several questions, the average value from the ratings was calculated. The ratings on these 15 capacities were subjected to a principal components factor analysis with varimax rotation (Kaiser normalization). The rotated solution yielded 3 factors with eigenvalues greater than 1 that explained 80% of the variance.

The first factor (*Factor 1*) accounted for 32% of the variance and included 7 capacities – *beliefs*, *conscious experience*, *agency*, *desires*, *planning*, *affective state*, *moral agency*. It seems that the first dimension is a combined Experience-Agency dimension.

The second factor (*Factor 2*) accounted for 26.7% of the variance and included 5 capacities – *personality*, *communication*, *self-control*, *theory of mind.*

The third factor (*Factor 3*) accounted for 21.2% of ratings variance and included 3 of the capacities – *cognitions, emotional experience, perception, psychobiological*.

The average ratings on each factor were calculated and subjected to a 3 x 3 Repeated-Measures ANOVA with *DMP* (*Factor1* vs. *Factor2* vs. *Factor3*) as a within-subjects factor and *identity of the agent* (*human* vs. *artificial human* vs. *humanoid robot*) as a between-subjects factor. The analysis revealed a main effect of DMP, $F(2, 114) = 10.52$, $p < .001$. A Bonferroni post-hoc test revealed that agents receive lower ratings ($M = 4.55$) on the second dimension than on the first dimension ($M = 5.25$, $p < .001$) and on the third dimension ($M = 5.31$, $p = .003$).

The main effect of *identity of the agent* is not statistically significant. The interaction is aslo not significant.

Moral Judgments

Proportion of the participants answering that the utilitarian action undertaken by the agent, is morally right is 0.5 for the human agent, 0.41 for the artificial human, 0.5 for the humanoid robot. Human, artificial human, and humanoid robot receive mean moral permissibility ratings of 3.1, 3.7, and 3.7 and blameworthiness ratings of 3.4, 3.3, and 3.0, respectively. No significant differences are found.

Summary of the Results in Experiment 2

In Experiment 2, again three dimensions of mind perception are revealed, but they are different from the dimensions identified in Experiment 2. The difference is attributed to the utilitarian moral action undertaken by the agent before the mind perception ratings being made. First dimension identified here combines mental capacities from Experience and Agency dimensions identified in Experiment 1.

No differences are found between agent's ratings on each of the identified dimensions in Experiment 2. It seems that undertaking the utilitarian moral action makes the artificial agents to be perceived as similar to the human agent.

Influence of Moral Action on Mind Perception

In order to explore further the influence of moral action on mind perception, we compared the ratings for each of the 15 mental capacities between Experiment 1 and Experiment 2. Description of the agent undertaking the utilitarian moral action preceded mind perception ratings in Experiment 2 so this is considered as *moral-action* condition.

Ratings for each mental capacity were analyzed in a 3x2 ANOVA with *identity of the agent* (human vs. artificial human vs. humanoid robot) and *agent's moral action* ('*no'* or '*yes'*) as between-subjects factors. Only the significant results are reported here.

Identity of the agent had an effect on the ratings of the following mental capacities: *conscious experience* ($p = .016$), *affective states* ($p = .002$), *emotional experience* ($p = .020$) and *desires* ($p < .001$). The *human* agent was rated higher than the *humanoid robot* on all of those mental capacities (all p's $< .05$). The human agent was rated higher than the

artificial human on *affective states* (p = .051) and *desires* (p = .017). For *beliefs* the effect was marginally significant (p = .054): *human agent* was rated higher than the *humanoid robot* (p = .065).

Agent's moral action had an effect on the ratings of the following mental capacities: *agency* (p = .014), *moral agency* (p = .029), *beliefs* (p = .024), *conscious experience* (p = .056), and *planning* (p = .058). The result is interesting, as it demonstrated that the agent's moral action have an effect not only on his agency and moral agency ratings, but also on the rating of other mental capacities.

Discussion and Conclusions

The paper investigates the dimensions of mind perception for human agents and fictitious artificial agents (an artificial human and a humanoid robot) that are identical to humans and how mind perception is affected by the agent being presented as moral agent.

In Experiment 1, three dimensions of mind perception are identified – Experience, Agency, and Cognition. The identified dimensions of mind perception are ascribed to different agents in a different pattern. The artificially created agents are judged higher on the Cognition dimension than on the Agency or Experience dimensions. The human is judged higher on the Experience and Cognition dimensions than on the Agency dimension. The artificial agents are rated as similar to the human agent on Agency and Cognition dimension but not on the Experience dimension.

People more readily ascribe cognitive mental capacities to artificially created agents than mental capacities belonging to the Experience or Agency dimensions.

In Experiment 2, the goal was to explore the influence of a utilitarian moral action undertaken by the agent on mind perception for that agent. The three dimensions of mind perception here are restructured – the first dimension regroups mental capacities that seem influenced by the preceding agents' moral action description like *agency, moral agency, consciousness, planning and affective* states. The second factor is related to communication and social interaction, while the third to cognition and psychobiological capacities. Now the artificial agents are rated to be similar to a human.

The results of the two experiments show that a utilitarian moral action undertaken by an agent has a strong effect no only on the evaluation of moral agency but also other mental capacities.

Another goal of the study was to explore the moral judgments about utilitarian moral action undertaken by those three agents. It turns out that there are no differences in the moral judgments for the human or the artificially created agents. This result is in line with the finding that similar agency and moral agency is ascribed to the human and to the artificial agents.

In conclusion, our results provide support for the idea that some mental states and capacities (especially cognitive ones) are more readily ascribed to non-human agents; while other mental states (related to conscious experience) are ascribed to a lesser extend to non-human agents. They also give evidence that mind perception space is sensitive to and dependent on the actions performed by an agent.

References

Arico, A., Fiala, B., Goldberg, R. F., & Nichols, S. (2011). The Folk Psychology of Consciousness. *Mind & Language*, *26*(3), 327–352.

Foot, P. (1967). The Problem of Abortion and the Doctrine of Double Effect. *Oxford Review*, *5*, 5–15.

Gray, H. M., Gray, K., & Wegner, D. M. (2007). Dimensions of mind perception. *Science*, *315*(5812), 619.

Heider, F., & Simmel, M. (1944). An experimental study of apparent behavior. *The American Journal of Psychology*, *57*(2), 243.

Hristova, E., & Grinberg, M. (2015). Should Robots Kill? Moral Judgments for Actions of Artificial Cognitive Agents. In *Proceedings of EAPS 2015*.

Hristova, E., Kadreva, V., & Grinberg, M. (2014). Moral Judgments and Emotions: Exploring the Role of 'Inevitability of Death'and "Instrumentality of Harm" (pp. 2381–2386). Austin, TX: Proceedings of the Annual Conference of the Cognitive Science Society.

Sparrow, R. (2007). Killer Robots. *Journal of Applied Philosophy*, *24*(1), 62–77.

Strait, M., Briggs, G., & Scheutz, M. (2013). Some correlates of agency ascription and emotional value and their effects on decision-making. In *Affective Computing and Intelligent Interaction,* 505-510. IEEE.

Sullins, J. (2006). When is a robot a moral agent? *International Review of Information Ethics, 6*, 23-30.

Takahashi, H., Terada, K., Morita, T., Suzuki, S., Haji, T., Kozima, H., et al. (2014). Different impressions of other agents obtained through social interaction uniquely modulate dorsal and ventral pathway activities in the social human brain. *Cortex*, *58*(C), 289–300.

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

Wallach, W., & Allen, C. (2008). *Moral Machines: Teaching Robots Right from Wrong*. Oxford University Press.

Waytz, A., Gray, K., Epley, N., & Wegner, D. M. (2010). Causes and consequences of mind perception. *Trends in Cognitive Sciences*, *14*(8), 383–388.

Ward, A. F., Olsen, A. S., & Wegner, D. M. (2013). The Harm-Made Mind: Observing Victimization Augments Attribution of Minds to Vegetative Patients, Robots, and the Dead. Psychological Science, 24(8), 1437–1445.