EMPATHY AND MORAL JUDGMENT IN TROLLEY-LIKE DILEMMAS

Veselina Kadreva, Evgeniya Hristova

Abstract: A large body of recent research in the field of moral psychology has established the role of emotional processing in judgment. Previous findings support the dual-process model of moral judgment [Greene et. al., 2001] suggesting that it is driven both by controlled cognitive processes and automatic emotional processes. The present study provides data in support of this model exploring emotional empathy and its association with moral judgment of emotionally salient (personal) moral dilemmas. Strictly controlled stimuli are used in order to explore this correlation. Overall, personal dilemmas are judged as less permissible and receive lower permissibility ratings than impersonal dilemmas, which confirms the importance of the previously established in the literature personal-impersonal distinction. Results also show that individuals who score low on a measure of trait emotional empathy towards people and animals give higher permissibility ratings for the utilitarian action in personal trolley-like moral dilemmas, e.g. lower scores on emotional empathy scale is correlated with higher permissibility ratings for highly emotional personal moral dilemmas. No such correlation is observed for impersonal dilemmas. This pattern of results adds to previous findings revealing the importance of the emotional component in moral judgment and underscores the specific role of empathy in judgment of personal moral dilemmas.

Keywords: moral judgment, empathy, moral dilemmas, emotions

ACM Classification Keywords: A.0 General Literature - Conference proceedings

Introduction

Moral judgments are often studied using moral dilemmas - descriptions of hypothetical situations that pose a conflict between moral rules. E.g., in the classical Trolley problem [Foot, 1978], one needs to decide whether it is morally permissible to pull a switch so that a trolley about to kill five people on its way is redirected to another track, killing one person only. An alternative situation is represented by the ‘Footbridge dilemma’ [Thomson, 1985]. Here, only pushing a stranger in front of it could stop the trolley headed towards the five people.

A lot of studies have established that people judge it morally permissible to pull the switch in the Trolley problem, but not to push the stranger in the Footbridge dilemma (e.g., [Greene et al., 2001, 2008];
This behavioral dissociation is interesting as, from utilitarian point of view, both situations are equal: the sacrifice of one person leads to saving five other. It appears that moral judgment cannot be explained only by the application of certain principles either complying with utilitarian or consequentialists ethical norms.

As trolley-like problems successfully capture this tension, representing a moral dilemma, they have been extensively used as stimuli in a large body of research in the field of moral psychology. What is more, this type of moral dilemmas allow for a lot of factors (e.g. severity of harm, personal interest, intentionality, etc.) to be manipulated so that their influence on judgment could be studied in a detailed manner.

**Emotions, empathy, and moral judgment**

Actually, the most prominent theory to explain this dissociation is the Dual-process theory introduced by Greene et al., 2001. It states that moral judgment is led both by controlled cognitive processes and automatic emotional processes. Controlled cognitive processes are in favor of utilitarian judgments: ‘as five people are more than one, doing harm for the greater good is permissible’. On the other hand, harm aversion produces an emotional response that, if strong enough, might interfere with rational utilitarian calculations and restrain from action. The authors claim that ‘personal’ infliction of harm in the ‘Footbridge dilemma’ produces a salient emotional response compared to the one in the ‘Trolley problem’ where harm is inflicted impersonally and this distinction in emotional reaction is responsible for the observed differences in judgment. Using a set of personal and impersonal moral dilemmas, they provide both response time and neuroimaging evidence in support of the theory.

A lot of further studies on emotion and moral judgment have also provided evidence in support of the dual-process model of moral judgment. E.g, [Valdesolo & Desteno, 2006] have established that preliminary induced positive mood might diminish the negative response produced by personal scenarios resulting in a greater proportion of utilitarian judgments. For impersonal dilemmas, no similar effect was observed. On the contrary, when negative affect was induced preceding judgment, participants judged moral transgressions as less permissible [Wheatley & Haidt, 2005]; [Eskine et al., 2011]. What is more, neurophysiological studies [Moretto et al., 2010] have established that judgment of personal dilemmas in neurotypical populations is accompanied by greater arousal that is related to the smaller proportion of utilitarian judgments. For patients with damaged ventromedial prefrontal cortex who failed to exhibit a visceral reaction to personal dilemmas, no similar behavioral pattern was observed.

Other studies on clinical groups with disrupted emotional processing have also demonstrated that diminished emotional response could explain a greater probability for endorsement of utilitarian judgments, specifically for personal dilemmas. E.g. patients with frontotemporal dementia, frontal
traumatic brain injury, clinical psychopaths and individuals with psychopathic tendencies who are characterized by deficits in emotionality, social affect and inability to produce an empathetic response, all exhibit abnormal patterns of moral judgment [Patil & Silani, 2014].

Further studies aiming to identify the key components of emotional processing have actually established that specificities in judgment are observed also for healthy adults, but could be explained by individual differences in predispositions to experience empathic response. In a sequence of three experiments, [Gleichgerrcht & Young 2013] provided consistent evidence that people who score low on empathic concern, as measured by IRI scale [Davis, 1983] are prone to judge in a utilitarian manner for personal dilemmas, but there is no specific difference for impersonal ones. What is more, controlling for gender, the authors established that even previously documented gender differences in judgment, might actually be attributed to gender differences in empathic response. These results are in line with dual-process theory: people low on empathic concern fail to produce an emotional response, strong enough to prevent them from utilitarian judgment for personal dilemmas which are considered emotionally salient and there is no difference for less emotional impersonal dilemmas.

**Factors Affecting Moral Judgment**

Apart from the well-documented distinction between personal and impersonal dilemmas, in the literature have been established three other important factors which need to be controlled for: Benefit Recipient (self vs. other benefit), Inevitability of death (avoidable vs. inevitable), and Instrumentality of harm (instrumental vs. incidental) [Christensen & Gomila, 2012].

Benefit Recipient: this factor reflects whether the suggested moral violation would lead to consequences beneficial for the protagonist or not. Research consistently show that people are more likely to approve of moral violations when consequences would bring benefit for themselves [Bloomfield, 2007]. What is more, self-beneficial dilemmas are rated as both more arousing and negative compared to other-beneficial dilemmas [Christensen et al., 2014].

Inevitability of death: the factor is also well-recognized in the literature: when the person to be sacrificed is going to die regardless of intervention (inevitable death), dilemmas are judged as morally permissible compared to dilemmas where the victim is not endangered (avoidable death) [Moore et al., 2008]. On the other hand, according to [Christensen et al., 2014] findings, avoidable and inevitable dilemmas are not discriminative in terms of arousal and valence.

Instrumentality of harm: Instrumental dilemmas are considered those in which the action is intentional, aiming to harm one person in order to save five other. In incidental dilemmas, harm is inflicted as a collateral damage, indirectly resulting from the intervention. When harm is inflicted as a side effect (incidental harm), it is judged as more permissible compared to intentionally induced harm (instrumental
dilemmas), (e.g. [Hauser, 2006]; [Mikhail, 2007], [Moore et al., 2008]) and no differences in valence or arousal are established [Christensen et al., 2014]. The results reviewed above demonstrate that these factors need to be accounted for and need to be strictly controlled in further studies on moral judgment.

**Goals and Hypotheses**

The present study aimed to study the on the association between moral judgment and trait empathy using the Bulgarian adaptation of the Emotional Empathy Tendency Scale (EETS) [Mehrabian & Epstein, 1972]. Based on previous research, personal and impersonal moral dilemmas were used, as this distinction is found to be important as moral judgments in personal moral dilemmas are considered to involve stronger emotions. Strictly controlled stimuli were used in order to study this relationship, exploring several other factors established as significant in judgment (inevitability of death and instrumentality of harm). In line with previous research, participants low on trait empathy were expected to be more prone to endorse utilitarian actions for personal dilemmas. No similar effect was expected for impersonal dilemmas.

**Method**

**Stimuli and Design**

In the current experiment, moral dilemmas in which three factors are varied, are used:

- Physical directness of harm - harm is inflicted by physical contact (personal harm) or is mediated through mechanical means (impersonal harm);
- Instrumentality of harm - harm is inflicted intentionally as an instrument to save other endangered people (instrumental harm) or is a byproduct of engagement in another activity, aiming to save more people threatened (incidental harm);
- Inevitability of death - deathful harm needs to be inflicted either to a person that is going to die anyway (inevitable death), or to a person that is not endangered by the situation described in the scenario (avoidable death).

Physical directness of harm (personal vs. impersonal), instrumentality of harm (instrumental vs. incidental) and inevitability of death (avoidable vs. inevitable) factors are manipulated in a within-subjects design. Each participant is presented with a total of 12 dilemmas.

Stimuli are designed based on 4 scenarios (2 avoidable scenarios and 2 inevitable scenarios) each with 3 possible resolutions - personal instrumental, impersonal instrumental and impersonal incidental resolutions. However, all personal dilemmas are designed so that harm is always inflicted as an
instrumental harm, so that did not allow for full-factorial design. As a result, 8 of the 12 dilemmas are instrumental (4 personal and 4 impersonal) and the other 4 dilemmas are impersonal incidental (see also Table 1).

Table 1. An example of avoidable dilemma with three possible resolutions.

<table>
<thead>
<tr>
<th>Scenario</th>
<th>Personal/Instrumental</th>
<th>Impersonal/Instrumental</th>
<th>Impersonal/Incidental</th>
</tr>
</thead>
<tbody>
<tr>
<td>You are in a factory. You are standing on a platform above a railway track. Some loaded trolleys are moving along the rails. One heavy loaded trolley is speeding towards five workers as its breaks had suddenly stopped working. There is no time for them to run away and they are going to die. The trolley could be stopped only if a heavy object is set on its way.</td>
<td>The only thing that you can do is to push the worker standing next to you on the platform. He is going to fall down on the rails. Together with the tools that he is equipped with, the worker is heavy enough to stop the moving trolley. He is going to die but the other five workers will be saved.</td>
<td>The only thing that you can do is to activate a control button and to release the safety belt of a worker hanging from a platform above the rails. The worker will fall onto the rails of the trolley. Together with the tools that he is equipped with, the worker is heavy enough to stop the moving trolley. He is going to die but the other five workers will be saved.</td>
<td>The only thing that you can do is to activate a control button and to release a large container hanging from a platform. It will fall onto the rails of the trolley. The container is heavy enough to stop the moving trolley. On the top of the container there is a worker who will also fall on the rails. He is going to die but the other five workers will be saved.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Question</th>
<th>Is it permissible to act as described?</th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th>Rating Scale</th>
</tr>
</thead>
<tbody>
<tr>
<td>To what extent is it permissible to act as described?</td>
</tr>
<tr>
<td>1 2 3 4 5 6 7</td>
</tr>
<tr>
<td>Forbidden Permissible Obligatory</td>
</tr>
</tbody>
</table>
All of the stimuli are constructed with the aim to strictly control for possible confounding factors identified in the previous research:

1) All of the stimuli are homogenously structured: introductory paragraph describes the situation, followed by one sentence that introduces the one and only means of escape; finally, a resolution is suggested in a third paragraph.

2) In all dilemmas, only two avoidable and two inevitable scenarios (introductory paragraph) are used. In order to manipulate physical directness of harm and instrumentality of harm only the resolution paragraphs are modified. Three versions of resolutions for each of the situations are designed: personal/instrumental, impersonal instrumental, and personal incidental. In such a way, we control for possible differences arising from the specific situations described.

3) In all situations a constant tradeoff between killing one person and saving five other persons is described.

4) In all dilemmas participants are assigned the role of the protagonist.

5) In all of the dilemmas there is no self-risk for the protagonist.

6) The introductory paragraph describes simply a presence of the protagonist in a certain working environment without explicitly assigning a specific role or any responsibilities to them.

7) All of the six endangered persons are identified with equal roles in the described working environment /one and the same for all six persons- workmen, miners, crew members/ thus suggesting equal responsibilities.

8) The endangered and potentially sacrificed persons are adults only.

9) All situations are designed to illustrate artificial scenarios in order to avoid potential confounding effects, e.g. familiarity with a certain situation or readily available personal opinion on resolutions.

Each situation is followed by one and the same question: ‘Is it permissible to act as described?’ with two possible responses – ‘Yes’ and ‘No’. After that, the described action is evaluated on a 7-point Likert scale (where ‘1’ stands for forbidden, ‘4’ stands for permissible, and ‘7’ stands for obligatory) answering the question ‘To what extent is it morally permissible to act as described?’.

For each dilemma the following measures were recorded:

- Number of responses ‘Permissible’: in further analyses, calculated as the average ratio of ‘Yes’ responses given to the question ‘Is it permissible to act as described?’.
- Permissibility ratings: participants indicated on a seven-point Likert scale, with anchors: 1 – forbidden, 4 – permissible, 7 – obligatory, whether the suggested action was morally permissible answering the question ‘To what extent is it morally permissible to act as described?’.
- Response time: time needed to give a response to the question ‘Is it permissible to act as described’ following a key press confirmation that they have read and understood the dilemma.
- Skin-conductance response (SCR) during the response period was recorded as a measure of autonomic arousal.

At the end of the experiment, participants filled out a paper-and-pencil version of EETS questionnaire [Stoyanova, 2011]. The EETS questionnaire is a self-report measurement used to evaluate empathy as a construct which reflects specifically the emotional response to perceived emotional experience of others, rather than cognitive components of empathy (e.g. perspective taking and theory of mind). High scores on the questionnaire indicate a high responsiveness to other peoples' emotional reactions.

The adapted version of the instrument [Stoyanova, 2011] consists of 25 items assessed on a 4-point response scale. Participants are presented with 25 statements describing certain behaviors and need to indicate on a 4 point response scale (from ‘1 – No, Never’ to ‘4 – Yes, Always’) to what extent each of the statements apply for themselves. Although, in the literature, total scores are reported, scores on 2 factors with high reliability are also subjected to further analyses. The two factors are ‘empathy towards people and animals’ (α = 0.76) and ‘empathy towards fictitious characters’ (α = 0.64). Sample items are presented at Table 2.

Table 2 Sample items of EETS.

<table>
<thead>
<tr>
<th>Item N</th>
<th>Sample Items</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>It makes me sad to see a lonely stranger in a group.</td>
</tr>
<tr>
<td>5</td>
<td>I tend to get emotionally involved with a friend’s problems.</td>
</tr>
<tr>
<td>7</td>
<td>I tend to lose control when I am bringing bad news to people.</td>
</tr>
<tr>
<td>6</td>
<td>Sometimes the words of a love song can move me deeply.</td>
</tr>
<tr>
<td>10</td>
<td>I like to watch people open presents.</td>
</tr>
<tr>
<td>20</td>
<td>I am very upset when I see an animal in pain.</td>
</tr>
</tbody>
</table>
Participants and Procedure

A total of 34 participants (11 males, 23 females) took part in the experiment. The age range was from 16 to 40 (M = 23). The participants took part in the experiment in exchange for partial credit toward an undergraduate course requirement or voluntarily.

Each participant was presented with 12 dilemmas – 4 situations (2 avoidable and 2 inevitable) each with 3 possible resolutions (personal instrumental, impersonal instrumental, and impersonal incidental). For each participant, the dilemmas were presented in a pseudo-randomized order ensuring that the same situation never appears in two consecutive dilemmas. In total, twelve pseudo-randomized orderings were used.

Participants were tested individually. First, the electrodes for recording skin conductance were put to provide enough time for the gel used to be absorbed. Next, the experimenter read the instructions. In the instructions, it was emphasized that participants had to imagine that the action described was the only action possible; that they had to disregard legality and had to consider only moral appropriateness of judgment. Each participant was asked to remain relatively still in order to avoid artifacts in the recordings.

First, four practice dilemmas were shown. Next, the twelve stimuli were presented using E-Prime 1.2 software. Each of the stimuli was presented on two consecutive screens: on screen 1 appeared only the first paragraph of the dilemma – the scenario. Participants indicated that they have read and understood the scenario by pressing the spacebar and advanced to the second screen which presented both the scenario and the resolution (Figure 1). Following a keypress confirmation for reading completion and understanding, participants advanced to the next screen, the dilemma text disappeared and the question: ‘Is it permissible to act as described?’ appeared on a new screen. Participants indicated either ‘Yes’ or ‘No’ using the computer keyboard and advanced to the final screen presenting a question: ‘To what extent is it permissible to act as described?’ and rating scale with anchors ‘1 – forbidden’, ‘4 – permissible’, ‘7 – obligatory’. Participants entered a number from 1-7 using the computer keyboard. Response was followed by 700 ms inter-trial interval (Figure 1).

After the participants completed the first part of the experiment, each of them filled out a paper-and-pencil version of EETS questionnaire adapted for Bulgarian population [Stoyanova, 2011].

Results

In the analyses, we used the responses for 8 of the 12 dilemmas. In those 8 dilemmas physical directness of harm and inevitability of death factors were varied resulting in 2 dilemmas for each combination of the factors’ levels. All of those 8 dilemmas were instrumental.
Responses Permissible

Number of responses ‘permissible’ was analyzed in a repeated measures ANOVA with physical directness of harm (personal vs. impersonal) and inevitability of death (avoidable vs. inevitable) as within-subjects factors. Analysis revealed main effect of physical directness of harm (F(1, 33 = 9.72), p = .004): impersonal harm was judged as more permissible than personal harm (57% vs. 41%) ‘permissible’ responses (Figure 2).

The interaction was not significant and there was no main effect of inevitability of death.

Permissibility ratings

Permissibility ratings (on a scale from ‘1 – forbidden’ to ‘7 – obligatory’) were analyzed in a repeated measures ANOVA with physical directness of harm (personal vs. impersonal) and inevitability of death (avoidable vs. inevitable) as within-subjects factors. Analysis revealed main effect of physical directness of harm (F(1, 33 = 8.42, p = .007): impersonal harm was judged as more permissible than personal harm (3 vs. 2.6) (Figure 3). The interaction was not significant and there was no main effect of inevitability of death.

Figure 1. A schematic representation of a trial.
Figure 2 Percentage of responses ‘permissible’ for each dilemma type.

Figure 3 Average permissibility ratings for each dilemma type.
Empathy (EETS) Total Scores and Moral Judgment

First, we explored the association between average scores on the EETS scale and average permissibility ratings for each of the participants using a Pearson correlation analysis. Higher scores on EETS questionnaire indicate a predisposition for high responsiveness to others’ emotional reactions and a tendency to experience a strong empathic response. No significant correlation was found between EETS scores and average permissibility ratings.

Then, average permissibility ratings for personal and impersonal dilemmas were entered in separate correlation analyses. For personal dilemmas, the correlation between EETS scores and average permissibility ratings was not statistically significant. For impersonal dilemmas, the correlation between EETS scores and average permissibility ratings was, also, not statistically significant.

Empathy (EETS Factor 1) scores and Moral Judgment

Next, we explored the association between average scores on the first factor of the EETS scale (‘Empathy towards people and animals’ subscale) and permissibility ratings. Higher scores on EETS factor 1 scale indicate a predisposition for high responsiveness to others’ emotional reactions, a tendency to experience a strong empathic response towards people and animals. First, average scores on EETS factor 1 (‘Empathy towards people and animals’) and average ‘permissibility’ ratings for all dilemmas for each of the participants were subjected to a Pearson correlation analysis. Analysis revealed that ‘permissibility’ ratings were negatively correlated with average factor 1 scores ($r(29) = - .37, p = .046$). Lower empathy scores on that factor are correlated with higher ratings about the permissibility of the utilitarian actions.

Then, the same correlation (between empathy EETS factor 1 scores and permissibility ratings) was studied in separate correlation analyses: one for personal dilemmas and one for the impersonal dilemmas.

For personal dilemmas, there was a significant negative correlation ($r(29)= -.40, p = .031$). For personal dilemmas, those who scored lower on factor 1, on average, gave higher permissibility ratings (Figure 4). For impersonal dilemmas, the correlation between EETS factor 1 scores and average permissibility ratings was not statistically significant.
Further, the association between factor 2 scores (empathy towards fictitious characters) and permissibility ratings was explored. Higher scores on this factor indicate a tendency for high emotional responsiveness to the perceived emotional reactions of fictitious characters.

The association between average factor 2 scores and permissibility ratings for each of the participants was explored using a Pearson correlation analysis. No significant correlation was found.

Then, average permissibility ratings for personal and impersonal dilemmas were entered in separate correlation analyses. For personal dilemmas, the correlation between EETS factor 2 scores and average permissibility ratings was not statistically significant. For impersonal dilemmas, the correlation between EETS factor 2 scores and average permissibility ratings was, also, not statistically significant.

**Discussion**

A large body of recent research have established that emotional processing is relevant in moral judgment. Prior studies on clinical groups with disrupted emotionality have revealed that these deficiencies might be related to and explain specific patterns in moral judgment. The present study adds to previous findings demonstrating that trait predispositions to exhibit emotional empathic response in
healthy participants are related to moral judgment as well: participants low on trait emotional empathy are characterized by patterns of judgment typical for clinical populations: they judge personal dilemmas as more permissible and there is no association between emotional empathy scores and permissibility ratings on impersonal dilemmas. The present findings are compliant with a lot of studies demonstrating that individuals inclined toward utilitarianism exhibit personality traits related to diminished emotional reactivity. These results are in line with Dual-process theory showing that diminished utilitarian judgment is related to susceptibility to dilemmas of high emotional salience.

Future work needs to explore in a more detailed manner associations with empathy as a multidimensional construct. Relationships between moral judgment and tendencies to experience personal distress, empathic concern, empathy towards fictitious characters and perspective taking (as a measure of cognitive empathy) need to be studied as well. Further research will also benefit from the inclusion of a self-benefit condition in order to establish more firmly the specific role of empathy as an ‘other-oriented’ emotion.

Acknowledgements

The paper is published with partial support by the ITHEA ISS (www.ithea.org), the ADUIS (www.aduis.com.ua). and the Central Fund for Strategic Development, New Bulgarian University.

Bibliography


**Authors’ Information**

**Veselina Kadreva** – a PhD candidate, Cognitive Science and Psychology Department, New Bulgarian University, New Bulgarian University, 21 Montevideo str., Sofia, Bulgaria; e-mail: vkadreva@nbu.bg

**Major Fields of Scientific Research**: Moral judgment, Biosignal-based research

**Evgeniya Hristova**, PhD – Research Center for Cognitive Science, Cognitive Science and Psychology Department, New Bulgarian University, 21 Montevideo str., Sofia, Bulgaria; e-mail: ehristova@cogs.nbu.bg

**Major Fields of Scientific Research**: Moral judgment, Social Dilemmas, Eye-tracking, Biosignal-based research