

Emotions associated with counterfactual comparisons drive decision-making in Footbridge-type moral dilemmas

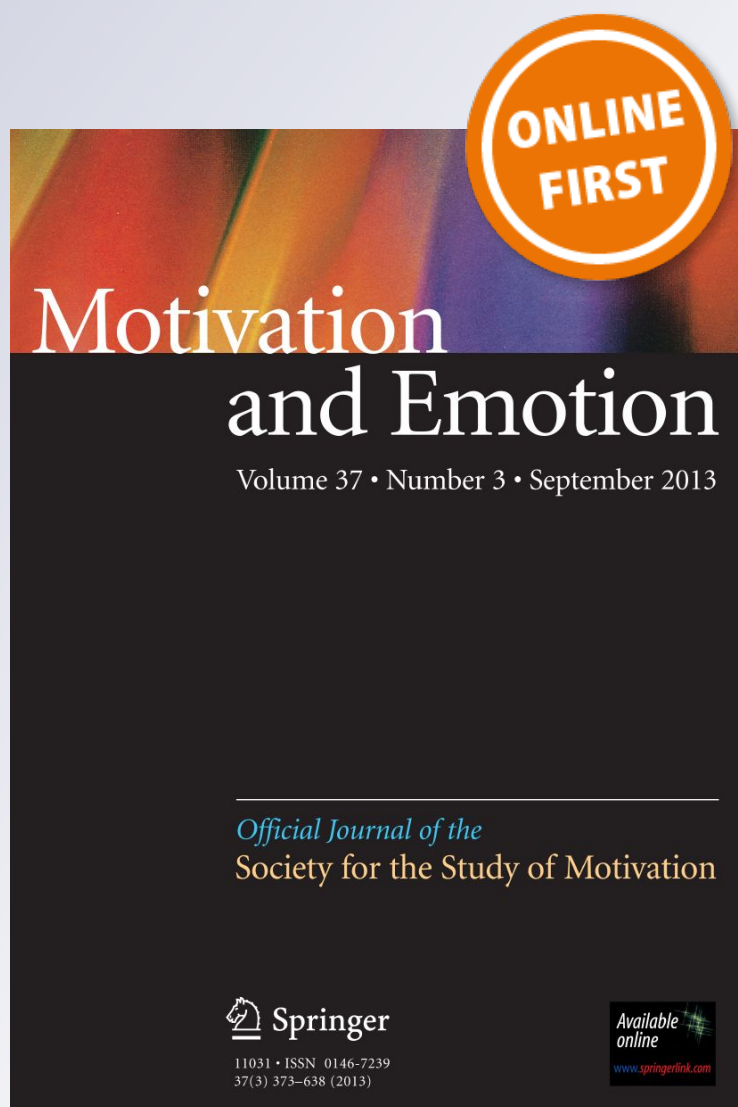
Alessandra Tasso, Michela Sarlo & Lorella Lotto

Motivation and Emotion

ISSN 0146-7239

Motiv Emot

DOI 10.1007/s11031-017-9607-9



Your article is protected by copyright and all rights are held exclusively by Springer Science +Business Media New York. This e-offprint is for personal use only and shall not be self-archived in electronic repositories. If you wish to self-archive your article, please use the accepted manuscript version for posting on your own website. You may further deposit the accepted manuscript version in any repository, provided it is only made publicly available 12 months after official publication or later and provided acknowledgement is given to the original source of publication and a link is inserted to the published article on Springer's website. The link must be accompanied by the following text: "The final publication is available at link.springer.com".

Emotions associated with counterfactual comparisons drive decision-making in Footbridge-type moral dilemmas

Alessandra Tasso¹ · Michela Sarlo^{2,4}  · Lorella Lotto^{3,4}

© Springer Science+Business Media New York 2017

Abstract Based on the dual-process theory of moral judgment, it has been suggested that in Footbridge-type dilemmas the anticipation of the emotional consequences of causing intentional harm might contribute to the decision of rejecting utilitarian resolutions. However, no empirical data have been reported on the emotions felt by participants after their decisions, and the role played by emotions in Trolley-type dilemmas remains to be determined. The present study investigated the specific emotions engaged both after decision choices and after the generation of the counterfactual scenario in Trolley- and Footbridge-type dilemmas. The results support the idea that in Footbridge-type dilemmas decision-making is driven by the attempt to minimize the aversive emotional state evoked by the decision outcome. A greater increase in emotional intensity was found overall for Footbridge-type than Trolley-type dilemmas after the counterfactual generation following typical (non-utilitarian) choices, with guilt, regret, and shame being the emotions that increased most. Critically, in Footbridge-type dilemmas only, typical choices were predicted by the increase in regret intensity experienced after counterfactual generation.

Keywords Moral dilemmas · Basic emotions · Moral emotions · Counterfactual · Decision-making

Introduction

Moral judgment can be broadly defined as the process by which people decide whether a course of action is right or wrong, including the evaluation of rights, duties, or obligations (e.g., Colby et al. 1980). While in some situations people can easily and unanimously judge the moral appropriateness of specific actions and behaviors, such as in intentionally hurting others, deciding is harder when conflicting moral principles or rights are at play. Indeed, particular set of circumstances, in which the rights of different individuals or stakeholder groups must be taken into account, may lead to divergent moral judgments. This is the case of moral *dilemmas*, typically involving a conflict in choosing between two undesirable alternatives, both of which have aversive consequences and none of which clearly emerges as the right thing to do (e.g., Braunack-Mayer 2001; Sinnott-Armstrong 1987), such as in choosing between killing one person and letting many people die. The Trolley and Footbridge problems are prototypical examples of this condition. In the Trolley dilemma, the only way to save five workmen from a runaway trolley is to pull a lever redirecting the trolley onto a sidetrack, where it will kill a single workman. In the Footbridge dilemma, the only way to save the five workmen is to push a large man off an overpass onto the track, where he will die while his body will stop the trolley. Despite the same cost/benefit ratio, moral judgments in the two dilemmas appear to be driven by different principles, as most people judge that pulling the lever in the Trolley dilemma is morally

✉ Michela Sarlo
michela.sarlo@unipd.it

¹ Department of Humanities, University of Ferrara, Ferrara, Italy

² Department of General Psychology, University of Padova, Via Venezia 8, 35131 Padova, Italy

³ Department of Developmental Psychology and Socialization, University of Padova, Padova, Italy

⁴ Center for Cognitive Neuroscience, University of Padova, Padova, Italy

acceptable, whereas pushing the man in the Footbridge dilemma is not (e.g., Hauser et al. 2007).

Empirical evidence suggests that intentionality of action plays a major role in producing divergent moral judgments. Indeed, people tend to choose action when harming others is a foreseen but *unintended* consequence of producing an overall greater good (as in the Trolley dilemma), while they prefer action omission when the harmful act is an *intended* means to pursue a greater good (as in the Footbridge dilemma) (Foot 1967; Cushman et al. 2006; Hauser et al. 2007). In addition, the negative emotions elicited by harm are found to increase judgments of moral blameworthiness only when the harm is intentional (Treadway et al. 2014).

According to the dual-process theory (Greene et al. 2004, 2001), emotion is the critical factor differentially affecting moral judgment in the two types of dilemmas. Specifically, based on different types of evidence, including behavioral, functional magnetic resonance, and event-related potential data (Borg et al. 2006; Greene et al. 2001, 2004; Moore et al. 2008; Sarlo et al. 2012), dilemma resolutions are hypothesized to be driven by the interaction between two competing processing systems mediated by partially dissociable neural networks: a fast, automatic emotional system engaging mainly the amygdala and the ventromedial prefrontal cortex (vmPFC), and a slow, controlled cognitive system engaging mainly the dorsolateral prefrontal cortex (DLPFC) and the inferior parietal lobe. In particular, cognitive processes would drive utilitarian choices (e.g., approving of killing one person to save more lives) in Trolley-type dilemmas, whereas emotional processes would prompt non-utilitarian choices (e.g., disapproving of killing one person to save more lives) in Footbridge-type dilemmas. Further evidence in support of the dual-process theory comes from neuropsychological studies that consistently showed an atypically high number of utilitarian responses to Footbridge-type dilemmas in patients with focal lesions to the vmPFC (Ciamarelli et al. 2007; Koenigs et al. 2007), suggesting a causal role played by emotional brain networks in rejecting utilitarian resolutions.

In spite of the large number of studies emphasizing the central role played by emotions in moral decision-making, to the best of our knowledge, only few studies have focused on the emotions felt by participants when responding to moral dilemmas. Specifically, Choe and Min (2011), using 25 Footbridge-type dilemmas (i.e., “personal” dilemmas in the Greene et al.’s 2001 study), asked participants to judge the appropriateness of the utilitarian resolutions and to choose the emotion most intensely felt during judgment from among anger, sadness, fear, anxiety, disgust, guilt, shame, surprise, and empathy. Results showed that guilt was the most intensely felt emotion, followed by sadness, disgust, anger, empathy, and anxiety. In another study by

Szekely and Miu (2015), where a subset of 12 Footbridge-type dilemmas was used, participants were first asked to choose between two courses of action (utilitarian/non-utilitarian) and then to identify the predominant emotion they were feeling during decision, by rating its intensity on a 5-point Likert scale. The results showed that fear and disgust were more frequently reported when participants rejected utilitarian resolutions, whereas regret was more frequently reported when participants accepted utilitarian resolutions. The frequency of the other emotions did not differ between utilitarian and non-utilitarian choices. Emotional intensity was overall higher for non-utilitarian than utilitarian choices.

Although these studies are valuable in highlighting the different emotions felt by participants *during* judgment or decision-making (Choe and Min 2011; Szekely and Miu 2015), they investigated Footbridge-type dilemmas only, providing no information about Trolley-type dilemmas. However, as there is a general agreement that both cognitive and emotional processes play a role in both Trolley- and Footbridge-type dilemmas (Greene et al. 2004; Gubbins and Byrne 2014; Manfrinati et al. 2013) it is of crucial interest to assess the emotions associated to both types of dilemmas in order to clarify their role in affecting decision choices.

According to the dual process theory of moral judgment (Greene et al. 2001, 2004), unpleasant emotions play a causal role in driving non-utilitarian choices in Footbridge-type dilemmas. This would imply that emotions are engaged more intensely during decision-making rather than after decision choice. In contrast, it might be argued that in Trolley-type dilemmas, emotions, while not playing any causal role in driving decisions, are strongly engaged by decision outcomes of utilitarian choices (i.e., the death of one person for the greater good).

Another relevant issue to be clarified in this context is the role played by the anticipation of future unpleasant emotional states in driving decision-making. Indeed, it has been suggested that in Footbridge-type dilemmas the anticipation of the emotional consequences of causing intentional harm might provide a substantial contribution to the decision of rejecting utilitarian resolutions (Ciamarelli et al. 2007; Sarlo et al. 2014). However, in the relevant literature no empirical studies have been reported to test this hypothesis.

On these bases, the first aim of the present study was to investigate in both Trolley- and Footbridge-type dilemmas the specific emotions engaged both after decision choices and after the generation of a counterfactual scenario. In particular, we aimed at comparing the two emotional patterns in terms of intensity and quality. To this end, we asked participants to rate the intensity with which they felt the six basic emotions (fear, anger,

disgust, sadness, joy, and surprise) and the main moral/counterfactual emotions (regret, disappointment, guilt, shame, and relief), both after the decision and after having imagined themselves in choosing the alternative resolution.

Traditionally, counterfactual thinking consists in the mental simulation of events, actions, or decisions, that is alternative to what happened in reality (e.g., Kahneman and Tversky 1982) and it is acknowledged that, when generating counterfactual scenarios, people experience a range of emotions such as regret, relief, blame, disappointment, and guilt (e.g., Gilovich and Medvec 1994; Johnson 1986; Kahneman and Miller 1986; Mellers et al. 1999). Importantly, decisions can arise from counterfactual comparisons in which people anticipate how they will feel about the different courses of action (see Loewenstein et al. 2001). Consistent with this view, in the field of moral decision-making, Sarlo et al. (2014) showed that in Footbridge-type, but not in Trolley-type dilemmas, the egoistic motivation to alleviate one's own distress (as measured by the Personal Distress subscale of the Interpersonal Reactivity Index; Davis 1980) drove the behavioral choices toward non-utilitarian resolutions. In other words, the authors suggest that the actual choice might be motivated by the counterfactual anticipation of the (unbearable) aversive emotional state associated with the (unchosen) utilitarian resolution. Although we did not directly measure anticipated emotions, the rationale of the paradigm employed in the present work was based on the assumption that the comparison between the emotions experienced after decision and after counterfactual may reflect the spontaneous (and possibly unconscious) comparison between the emotional costs of the two alternative resolutions people might make (during decision-making) before choice. We hypothesized that in Footbridge-type, but not in Trolley-type dilemmas negative emotions would be overall more intensely felt after the counterfactual generation, due to the greater emotional cost associated with the outcome of the unchosen option.

Our second research aim was to demonstrate a causal relationship between the negative emotions felt after counterfactual generation (especially guilt and regret) and decision choice in Footbridge-type dilemmas. Based on the dual process theory of Greene et al. (2001, 2004) and the results by Sarlo et al. (2014), we hypothesized counterfactual/moral emotions to predict the rejection of utilitarian decisions in Footbridge-type but not in Trolley-type dilemmas. Whether during the resolution of Trolley-type dilemmas counterfactual/moral emotions play a role in shaping decision choices or act as a mere correlate of decision-making will be also clarified.

Methods

Participants

A total of 148 undergraduate students (121 females, age range 18–29 years) at the University of Ferrara voluntarily participated in the study. Because of too many missing responses (>65%), the data of 17 participants were dropped from the analyses. The final sample, then, consisted of 131 participants (111 females).

Participants were tested in a classroom setting, they were informed that their responses would remain anonymous and gave written consent before participation. The study was conducted according to the Declaration of Helsinki and according to the deontological rules of the Italian Psychological Association (AIP).

Stimulus material

Participants received a booklet containing the instructions, and the texts of 20 moral and 5 filler dilemmas. All dilemmas were drawn from the standardized set of Lotto, Manfrinati, and Sarlo (2014). Detailed information on the criteria used to develop this set of stimuli can be found in Sarlo et al. (2012) and in Lotto et al. (2014). We employed 10 Footbridge-type dilemmas, which described killing one individual as an intended means to save others, and 10 Trolley-type dilemmas, which described killing one individual as a foreseen but unintended consequence of saving others. The 5 filler dilemmas involved no deaths and described other moral issues, such as lying or being dishonest (for examples, see Table 1).

The moral dilemmas were selected from the normative dataset of Lotto et al. (2014) on the basis of the following criteria: (a) we selected only the dilemmas in which the agent's life was not in danger, thus excluding the dilemmas in which killing one individual resulted in saving one's own and other people's lives ("self-involvement dilemmas"), (b) we selected the 10 Trolley-type dilemmas that had received the highest percentages of utilitarian responses ($M = 70.84\%$, $SD = 6.91$, range = 61.67–81.67%), (c) we selected the 10 Footbridge-type dilemmas that had received the lowest percentages of utilitarian responses ($M = 12.58\%$, $SD = 4.77$, range = 7.50–24.17%). Footbridge- and Trolley-type dilemmas were matched for the number of numerical consequences (i.e., the number of people to save or to let die).

Procedure

Each participant was presented with a 25 pages booklet. On each page there was the text of one moral dilemma. Each scenario ended with a sentence describing the proposed

Table 1 Sample Trolley-type, Footbridge-type and Filler dilemmas (text translated from Italian)

Dilemma	Scenario	Utilitarian resolution	Possible counterfactual
Trolley	You are in charge of a work crew who are doing repair work for the railways. In the distance you see a trolley and realize that the driver has lost control of it. If the trolley continues on it will end up running into five workers who are working on the tracks. On a secondary track there is one worker	You pull a lever on the interchange which will divert the trolley onto the secondary track. You know that it will run into and kill the worker, but the other workers will be unharmed	If I had decided _____ [not to pull the lever, the 5 workers would have died]
Footbridge	You are crossing a bridge which passes over the railway line and you see a train moving on the track below. You notice that the driver has lost control and that the locomotive will continue its course and end up running into five men who are working on the adjacent tracks. On the bridge next to you is a person you do not know	You push this person off the bridge is such a way that his body blocks the oncoming train. You know that his person will die, but the five workers will be unharmed	If I had decided _____ [not to push the man off the bridge, the 5 workers would have died]
Filler	Due to the last year's economic crisis, the company in which you worked became bankrupt and you lost your job. Lately, you have been looking for a new job, but unsuccessfully. You've realized you lack competence in informatics and you believe you would be hired more easily if you had a CV with such characteristics	You include in your CV some false information about competence in informatics. You will overcome other candidates who are more skilled than you and will be hired	If I had decided _____ [not to include false information in the CV, I would not have obtained the new job]

Possible counterfactuals generated by participants after choosing the utilitarian resolution are reported in squared brackets

utilitarian resolution (see Table 1), and participants were asked to indicate whether they would perform the proposed action by choosing between “yes” and “no”.

After their decision, participants rated their actual emotional state by indicating the intensity with which they felt the six basic emotions (fear, anger, disgust, sadness, joy, and surprise) and the main moral/counterfactual emotions (regret, disappointment, guilt, shame, and relief). Following Giorgetta, Zeelenberg, Ferlazzo and D'Olimpio (2012), three different Italian words were used for regret (“rimorso” = *action regret*, “rimpianto” = *inaction regret*, and “rammarico” = *general regret*) and two for disappointment (“disappunto” = *general disappointment*, and “delusione” = *action/inaction disappointment*)¹. The intensity of each emotion was evaluated on a 5-point Likert-type scale, where 0 meant “I don't feel this emotion at all” and 4 meant “I feel this emotion very intensely”. The task was introduced by the sentence “How do you feel now, after the decision? To what extent do you feel the following emotions?”

Then, participants were asked to imagine to choose the alternative (unchosen) resolution and to complete the conditional sentence “If I had decided to _____” and to describe the counterfactual consequences of the alternative course of action (e.g., “If I had decided to *push the man off the bridge, the 5 workers would be alive*”). This task was introduced by the sentence “Imagine that you had chosen the alternative resolution. What would have been the consequences of this choice?”

After completing the conditional sentence, participants were asked to rate again their current emotional state using the same 5-point Likert-type scale, with the 14 emotions presented in a different order. This task was introduced by the following sentence “How do you feel now, after having imagined that you had chosen the alternative resolution? To what extent do you feel the following emotions?”

At the conclusion of the experiment, participants were debriefed, thanked, and dismissed.

Data analysis

An analysis of variance (ANOVA) was carried out in order to compare the percentages of utilitarian responses in the

¹ The different meanings of the specific words depend on their association with responsibility. Specifically, in English, “regret” is associated to both action and inaction, whereas in Italian “rimorso” is associated to regret for action, “rimpianto” is associated to regret for inaction, and “rammarico” refers to a general negative feeling following a negative outcome. In a similar way, the Italian word “disappunto” corresponds to an attribution-independent emotion, whereas “delusione” is associated with the agent's responsibility, that is, both with action and inaction.

two types of dilemmas, with *Type of Dilemma* as a within subject factor (Trolley-type vs. Footbridge-type).

A second ANOVA was conducted on the mean intensities of the emotions, with *Moral Condition* (actual decision vs. counterfactual generation), *Type of Dilemma* (Trolley-type vs. Footbridge-type), and *Emotion* (anger, disgust, fear, sadness, joy, surprise, action regret, inaction regret, general regret, action/inaction disappointment, general disappointment, guilt, shame, relief) as within-subject factors.

Two separate regression analyses were run for Trolley-type and Footbridge-type dilemmas on the percentages of typical choices, using as predictor variables the difference scores between the emotion intensities after counterfactuals and those after decision choices.

In order to focus on the processes involved in the typical dilemma resolutions, only the typical responses, i.e., utilitarian responses for Trolley-type dilemmas and non-utilitarian responses for Footbridge-type dilemmas, were considered ($N = 123^2$, 106 females).

The corrected p -values for effects within variables with more than two levels are reported together with the Greenhouse-Geisser epsilon (ϵ) and the uncorrected degrees of freedom. The Bonferroni-corrected post-hoc comparisons were conducted on significant main effects and interactions.

All analyses were performed using IBM SPSS Statistics, 21 (Armonk, New York, USA).

Results

Percentages of utilitarian responses

As expected, Trolley-type dilemmas elicited a higher proportion of utilitarian responses than Footbridge-type dilemmas [$F(1,130) = 372.32$, $p < .0001$, $\eta_p^2 = 0.74$; $M_s = 64.20\%$ vs. 15.58% , respectively].

Emotional state

The ANOVA showed a significant main effect of *Moral Condition* [$F(1,122) = 136.68$, $p < .0001$, $\eta_p^2 = 0.53$], with more intense emotions experienced after counterfactual than after decision ($M_s = 2.37$ vs. 1.92), a main effect of *Type of Dilemma* [$F(1,122) = 30.70$, $p < .0001$, $\eta_p^2 = 0.20$], with Trolley-type dilemmas eliciting overall more intense emotions than Footbridge-type dilemmas ($M_s = 2.23$ vs. 2.05 , respectively), and a main effect of *Emotion* [$F(13,1586) = 254.52$, $p < .0001$, $\epsilon = 0.55$, $\eta_p^2 = 0.68$].

² Eight additional subjects were excluded by the statistical program because they did not provide typical responses for any of the two dilemma types.

Table 2 Mean emotion intensities (and standard deviations) experienced after actual decision and after counterfactual generation for typical responses to Trolley-type and Footbridge-type dilemmas

Emotion	Trolley-type		Footbridge-type		Pairwise comparison			
	Decision (action) <i>a</i>	Counterfactual (no action) <i>b</i>	Decision (no action) <i>c</i>	Counterfactual (action) <i>d</i>	<i>a</i> vs. <i>b</i>	<i>c</i> vs. <i>d</i>	<i>a</i> vs. <i>c</i>	<i>b</i> vs. <i>d</i>
Sadness	3.16 (0.09)	3.40 (0.08)	3.06 (0.08)	3.45 (0.08)	0.0001	0.0001	0.13	0.39
Guilt	3.15 (0.07)	3.38 (0.08)	2.12 (0.10)	3.60 (0.06)	0.0001	0.0001	0.0001	0.0001
Anger	2.91 (0.10)	3.16 (0.09)	2.62 (0.10)	3.13 (0.09)	0.0001	0.0001	0.001	0.55
Fear	2.66 (0.11)	2.70 (0.11)	2.80 (0.11)	2.84 (0.11)	0.50	0.55	0.07	0.007
Action regret	2.52 (0.09)	2.98 (0.09)	1.81 (0.10)	2.97 (0.09)	0.0001	0.0001	0.0001	0.92
General regret	2.50 (0.09)	2.64 (0.10)	1.88 (0.10)	2.63 (0.10)	0.039	0.0001	0.0001	0.86
Shame	2.36 (0.10)	2.70 (0.11)	1.37 (0.11)	3.03 (0.10)	0.0001	0.0001	0.0001	0.0001
Action/inaction disappointment	2.34 (0.11)	2.72 (0.11)	1.91 (0.11)	2.70 (0.11)	0.0001	0.0001	0.0001	0.66
Inaction regret	2.28 (0.10)	2.89 (0.09)	1.64 (0.10)	2.87 (0.10)	0.0001	0.0001	0.0001	0.78
Disgust	2.09 (0.11)	2.43 (0.12)	1.77 (0.11)	2.76 (0.11)	0.0001	0.0001	0.003	0.0001
General disappointment	2.05 (0.11)	2.34 (0.11)	1.67 (0.11)	2.41 (0.11)	0.0001	0.0001	0.0001	0.20
Surprise	0.76 (0.10)	0.73 (0.10)	0.73 (0.09)	0.80 (0.10)	0.64	0.36	0.57	0.13
Relief	0.73 (0.08)	0.32 (0.05)	0.33 (0.05)	0.34 (0.05)	0.0001	0.92	0.0001	0.69
Joy	0.31 (0.05)	0.15 (0.04)	0.13 (0.03)	0.16 (0.04)	0.006	0.47	0.0001	0.91

p values of pairwise comparisons within (*a* vs. *b* and *c* vs. *d*) and between (*a* vs. *c* and *b* vs. *d*) dilemma types are also reported

The *Moral Condition* × *Type of Dilemma* interaction was significant [$F(1,122)=75.63$, $p < .0001$, $\eta_p^2 = 0.38$], showing that the increase in emotional intensity after the counterfactual generation was larger for Footbridge- than Trolley-type dilemmas. Moreover, the emotional intensity was larger for Trolley- than Footbridge-type dilemmas after decision ($p = .0001$; $M_s = 2.13$ vs. 1.70 , respectively), whereas it was larger for Footbridge- than Trolley-type dilemmas after counterfactual ($p = .003$; $M_s = 2.41$ vs. 2.32 , respectively).

The other two-way interactions were significant (all $p_s < .0001$) and were qualified by the significant *Moral Condition* × *Type of Dilemma* × *Emotion* interaction [$F(13,1586)=21.91$, $p < .0001$, $\epsilon = 0.47$, $\eta_p^2 = 0.15$]. Mean emotion intensities as a function of type of dilemma and moral condition are reported in Table 2 along with *p* values of pairwise comparisons within and between Trolley-type and Footbridge-type dilemmas.

In particular, for both Footbridge- and Trolley-type dilemmas all negative emotions (except fear) were more intensely felt after the counterfactual generation than after the actual decision. Shame, guilt, inaction regret and action regret were the emotions that increased most in the Footbridge-type dilemmas, whereas inaction regret, action regret, and action/inaction disappointment were those that increased most in the Trolley-type dilemmas. However, the intensity of the positive emotions, namely, relief and joy, was lower in the counterfactual than in the decision condition for Trolley-type dilemmas, whereas no significant

differences were observed for Footbridge-type dilemmas in the two moral conditions.

Importantly, results also showed that when participants gave utilitarian responses to Trolley-type dilemmas, sadness, guilt, and anger were the emotions most intensely felt both after decision and after counterfactual generation. In contrast, when participants gave non-utilitarian responses to Footbridge-type dilemmas, sadness, anger and fear were the emotions most intensely felt after decision, whereas sadness, anger and *guilt* were the emotions that prevailed after counterfactual generation.

Regression analyses

Two separate regression analyses were run in order to test whether the decision choices were predicted by the difference in intensity between the emotions experienced after decision and after counterfactual generation (i.e., the net increase in emotional cost resulting from the comparison between the two alternatives). Specifically, for each type of dilemma, a multiple regression analysis was run on the percentages of typical choices, using as predictor variables the difference scores between the emotion intensities after counterfactuals and those after decision choices. Tests for multicollinearity among predictors indicated that a low level of multicollinearity was present (tolerance values ranged from 0.19 to 0.62).

For Footbridge-type dilemmas, the regression model was significant [$F(11,126)=2.38$, $p = .011$, $R^2 = 0.19$].

Interestingly, only action regret (one of the emotions that increased most) proved to be a significant predictor of non-utilitarian responses ($\beta=0.59$, $B=11.69$, $SE(B)=3.82$, $t=3.06$, $p=.003$). For Trolley-type dilemmas, the model was not significant ($p=.44$).

Discussion

In the present study, we aimed at assessing the role of emotion in shaping decision-making in moral dilemmas. In particular, we investigated the specific emotions engaged both after decision choices and after the generation of a counterfactual scenario in both Trolley-type and Footbridge-type dilemmas.

Overall, the results showed that when faced with Trolley-type dilemmas participants reported sadness, anger, and guilt as the emotions most intensely felt both after the actual decision and after the counterfactual generation. In contrast, for Footbridge-type dilemmas, sadness, anger and fear were reported as the emotions most intensely felt after the actual decision, whereas sadness, anger, and guilt (instead of fear) were the emotions most intensely felt after the counterfactual. These results complement the findings reported by Choe and Min (2011) who indeed showed that guilt was the emotion most intensely felt *during* judgment in Footbridge-type dilemmas.

Our main hypothesis was that in Footbridge-type, but not in Trolley-type dilemmas, negative emotions would have been overall more intensely felt after the counterfactual generation. Results showed that, despite this effect was found for both types of dilemmas, the emotional intensity increased to a greater extent in Footbridge-type than in Trolley-type dilemmas. We believe that several results of the present study support the idea that in Footbridge-type dilemmas decision-making is driven by the attempt to minimize the unpleasant emotional state evoked by the decision outcome. First, a greater increase in emotional intensity was found overall for Footbridge- than Trolley-type dilemmas after the counterfactual generation, with the moral emotions of guilt, regret, and shame being the emotions that increased most. Importantly, in Footbridge-type dilemmas guilt emerged among the emotions most intensely felt only after counterfactual generation. Second, for most unpleasant emotions, emotional intensity after decision was significantly lower in Footbridge- than in Trolley-type dilemmas. Third, and most critically, typical choices were significantly predicted by emotion intensity in Footbridge-type dilemmas only. In particular, the increase in action regret intensity positively predicted non-utilitarian choices, indicating that the higher the emotional cost of the counterfactual alternatives (i.e., utilitarian choices), the higher the number of non-utilitarian choices. It is worth noting that it

is the differential increase, rather than the absolute emotional cost associated with the counterfactual alternative that seems to affect decision choices. We therefore suggest that in Footbridge-type dilemmas the comparison between the emotional costs ascribed to the outcomes of the available options strongly drives decision-making. In particular, we argue that in Footbridge-type dilemmas emotional intensity experienced during decision-making is reduced by the decision to choose non-utilitarian resolutions. Our findings strengthen and extend the dual process model of moral judgment (Greene et al. 2001, 2004) in showing that emotions play a critical and possibly causal role in driving decision-making toward non-utilitarian resolutions, in Footbridge-type but not in Trolley-type dilemmas. Specifically, we argue that the non-utilitarian resolution is chosen as the only option left after the rejection of the alternative (i.e., utilitarian) option, which plausibly poses an unbearable emotional burden.

Our data also provide important and novel insights on the role played by emotions in Trolley-type dilemmas. While it has been demonstrated that emotional processes are involved in this kind of dilemmas (Lotto et al. 2014; Manfrinati et al. 2013; Sarlo et al. 2012), their role in relation to decision-making has not been investigated in detail. According to Greene (e.g., 2008), the nature of the actions represented in Trolley-type dilemmas fails to trigger an emotional response comparable to that elicited in Footbridge-type dilemmas. Emotions in Trolley-type dilemmas have been often considered in the literature as processes too low in intensity to override the rational cost/benefit analysis (e.g., Greene 2008). However, some studies failed to find any difference between Trolley-type and Footbridge-type dilemmas in unpleasantness (Lotto et al. 2014; Pletti et al. 2015) or arousal (Lotto et al. 2014; Manfrinati et al. 2013; Pletti et al. 2015; Sarlo et al. 2012) experienced *during* decision-making. In the present study, the comparison between the emotions experienced *after* the actual decision and *after* the counterfactual generation shed light on the relationship between emotion and decision processes in Trolley-type dilemmas.

As mentioned above, when participants gave utilitarian responses to Trolley-type dilemmas, sadness, guilt, and anger were the emotions most intensely felt *both after decision and after counterfactual generation*, indicating that the pattern associated with the two alternatives was qualitatively comparable. Furthermore, although the emotional intensity was overall greater after the counterfactual generation than after the actual decision, such difference was significantly lower as compared with Footbridge-type dilemmas. Importantly, and differently from Footbridge-type dilemmas, the change in emotional intensity resulting from the comparison between the two alternatives did not show any relationship with decision choice. On these bases,

we suggest that in Trolley-type dilemmas emotion does not effectively contribute to decision-making because it cannot provide critical information about the different emotional costs of the two alternatives. As a consequence, the cognitive evaluation of harms and benefits prevails, leading to act in a utilitarian way. We therefore propose that in Trolley-type dilemmas emotions mainly represent a correlate of the conflict experienced as a consequence of the dilemmatic situation, which is highly unpleasant by definition (e.g., Braunack-Mayer 2001; Sarlo et al. 2012; Sinnott-Armstrong 1987), involving anger, sadness, and guilt. In this context, sadness might be elicited by the feeling of being the source of another person's pain or physical harm; anger might deal with the frustration of being forced to choose between undesirable alternatives and/or being unable to actively stop an offensive behavior (e.g., Hutcherson and Gross 2011); guilt might be elicited by the feeling of direct or indirect responsibility for somebody's harm, including failing to prevent harm (Prinz and Nichols 2010). Another relevant finding on Trolley-type dilemmas concerns the intensity of the pleasant emotions, i.e. joy and relief, whose investigation has been understandably neglected in the context of moral dilemmas. Our data showed that the intensity of these two emotions was greater after the actual decision than after the counterfactual generation for Trolley-type dilemmas while it did not differ for Footbridge-type dilemmas. This effect is plausibly related to the consequence of the choice of saving a greater number of people.

Taken together, our findings suggest that when harming others is an *unintended* consequence of producing an overall greater good, such as in the Trolley-type dilemmas, the outcomes of the two alternative courses of action is qualitatively similar in terms of emotional consequences, with a small difference in emotional intensity. The emotional pattern emerging after decision-making seems to confirm the cognitive-based decision, in that the emotional correlates of the utilitarian resolution, originating from a rational cost/benefit analysis, include lower negative and higher positive emotions as compared with the alternative resolution. In contrast, when harming others is an *intended* means to pursue a greater good, such as in the Footbridge-type dilemmas, the emotional pattern associated with the outcome of the two alternatives is different, both qualitatively and quantitatively. In our opinion, it is this differential information provided by the emotional system that allows rejecting the utilitarian option, as demonstrated by the finding that the change in intensity of action regret (one of the emotions that increased most) proved to be a significant predictor of non-utilitarian responses.

Some limitations of the present study should be acknowledged. First, the current research employed only self-report measures of emotion. Self-report measures are known to reflect an approximation of subjective

experience (e.g., Larsen and Fredrickson 1999) and to provide information only about those emotional states that reach awareness. Therefore, it is possible that in this context unconscious emotional processes play a role that we could not assess. Along the same line, despite decisions being often based on predictions about how the different options will make us feel, our ability to forecast our own feelings is not always accurate. In particular, while predictions about valence and specific emotions appear to be quite accurate, the intensity and the duration of future emotional states can be easily overestimated (Wilson and Gilbert 2003). Furthermore, as mentioned above, the affective encoding of the consequences of alternative courses of action may also operate implicitly (e.g., Damasio 1994). Lastly, while this study was not designed to examine the emotional impact of the choice per se (i.e., action/no action), by analyzing only the typical responses we were not able to disentangle the emotional effects of dilemma type from those associated with decision choice.

As a final remark, we believe that this kind of hypothetical scenarios, which may seem extreme and unrealistic, reflects ethical issues relevant to the challenges we are going to face in the near future. As an example, consider the mass-production of fully autonomous vehicles, which will likely be available on the market in the next few years. Although designed to reduce traffic accidents, these cars must be programmed to deal with unlikely but possible moral dilemmas, such as choosing whether to sacrifice their passengers to minimize harm (Bonnefon et al. 2016; Greene 2016). In this light, it seems crucial and timely to understand the complex cognitive-emotional interactions underlying such moral decisions, as we are far from identifying a set of sharable rules to be implemented in a "moral algorithm".

References

- Bonnefon, J. F., Shariff, A., & Rahwan, I. (2016). The social dilemma of autonomous vehicles. *Science*, 352, 1573–1576.
- Borg, J. S., Hynes, C., Van Horn, J., Grafton, S., & Sinnott-Armstrong, W. (2006). Consequences, action, and intention as factors in moral judgments: An fMRI investigation. *Journal of Cognitive Neuroscience*, 18, 803–817.
- Braunack-Mayer, A. J. (2001). What makes a problem an ethical problem? An empirical perspective on the nature of ethical problems in general practice. *Journal of Medical Ethics*, 27, 98–103.
- Choe, S. Y., & Min, K. H. (2011). Who makes utilitarian judgments? The influences of emotions on utilitarian judgments. *Judgment and Decision Making*, 6, 580–592.
- Ciaramelli, E., Muccioli, M., Ladavas, E., & di Pellegrino, G. (2007). Selective deficit in personal moral judgment following damage to ventromedial prefrontal cortex. *Social Cognitive and Affective Neuroscience*, 2, 84–92.

- Colby, A., Gibbs, J., Kohlberg, L., Speicher-Dubin, B., & Candee, D. (1980). *The measurement of moral judgment* (Vol. I). Cambridge, MA: Center for Moral Education, Harvard University.
- Cushman, F., Young, L., & Hauser, M. (2006). The role of conscious reasoning and intuition in moral judgment: Testing three principles of harm. *Psychological Science, 17*, 1082–1089.
- Damasio, A. R. (1994). *Descartes' error: Emotion, reason, and the human brain*. New York: Putnam.
- Davis, M. H. (1980). A multidimensional approach to individual differences in empathy. *JSAS Catalog of Selected Documents in Psychology, 10*, 85.
- Foot, P. (1967). The problem of abortion and the doctrine of double effect. *Oxford Review, 5*, 5–15.
- Gilovich, T., & Medvec, V. H. (1994). The temporal pattern to the experience of regret. *Journal of Personality and Social Psychology, 67*, 357–365.
- Giorgetta, C., Zeelenberg, M., Ferlazzo, F., & D'Olimpio, F. (2012). Cultural variation in the role of responsibility in regret and disappointment: The Italian case. *Journal of Economic Psychology, 33*, 726–737.
- Greene, J. D. (2008). The secret joke of Kant's soul. In W. Sinnott-Armstrong (Ed.), *Moral psychology. Vol. 3: The neuroscience of morality: Emotion, brain disorders, and development* (pp. 35–79). Cambridge, MA: MIT Press.
- Greene, J. D. (2016). Our driverless dilemma. *Science, 352*, 1514–1515.
- Greene, J. D., Nystrom, L. E., Engell, A. D., Darley, J. M., & Cohen, J. D. (2004). The neural bases of cognitive conflict and control in moral judgment. *Neuron, 44*, 389–400.
- Greene, J. D., Sommerville, R. B., Nystrom, L. E., Darley, J. M., & Cohen, J. D. (2001). An fMRI investigation of emotional engagement in moral judgment. *Science, 293*, 2105–2108.
- Gubbins, E., & Byrne, R. M. (2014). Dual processes of emotion and reason in judgments about moral dilemmas. *Thinking & Reasoning, 20*, 245–268.
- Hauser, M., Cushman, F., Young, L., Jin, R.K.X., & Mikhail, J. (2007). A dissociation between moral judgments and justifications. *Mind & Language, 22*, 1–21.
- Hutcherson, C. A., & Gross, J. J. (2011). The moral emotions: A social-functional account of anger, disgust, and contempt. *Journal of Personality and Social Psychology, 100*, 719–737.
- Johnson, J. T. (1986). The Knowledge of What Might have Been Affective and Attributional Consequences of Near Outcomes. *Personality and Social Psychology Bulletin, 12*, 51–62.
- Kahneman, D., & Miller, D. T. (1986). Norm theory: Comparing reality to its alternatives. *Psychological Review, 93*, 136.
- Kahneman, D., & Tversky, A. (1982). The simulation heuristic. In D. Kahneman, P. Slovic & A. Tversky (Eds.), *Judgment under uncertainty: Heuristics and biases* (pp. 201–208). Cambridge, MA: Cambridge University Press.
- Koenigs, M., Young, L., Adolphs, R., Tranel, D., Cushman, F., Hauser, M., & Damasio, A. (2007). Damage to the prefrontal cortex increases utilitarian moral judgements. *Nature, 446*, 908–911.
- Larsen, R. J., & Fredrickson, B. L. (1999). Measurement issues in emotion research. In D. Kahneman, E. Diener & N. Schwarz (Eds.), *Well-being: Foundations of hedonic psychology* (pp. 40–60). New York, NY: Russell Sage.
- Loewenstein, G. F., Weber, E. U., Hsee, C. K., & Welch, N. (2001). Risk as feelings. *Psychological Bulletin, 127*, 267–286.
- Lotto, L., Manfrinati, A., & Sarlo, M. (2014). A new set of moral dilemmas: Norms for moral acceptability, decision times, and emotional salience. *Journal of Behavioral Decision Making, 27*, 57–65.
- Manfrinati, A., Lotto, L., Sarlo, M., Palomba, D., & Rumiati, R. (2013). Moral dilemmas and moral principles: When emotion and cognition unite. *Cognition & Emotion, 27*, 1276–1291.
- Mellers, B., Schwartz, A., & Ritov, I. (1999). Emotion-based choice. *Journal of Experimental Psychology: General, 128*, 332–345.
- Moore, A. B., Clark, B. A., & Kane, M. J. (2008). Who shalt not kill? Individual differences in working memory capacity, executive control, and moral judgment. *Psychological Science, 19*, 549–557.
- Pletti, C., Sarlo, M., Palomba, D., Rumiati, R., & Lotto, L. (2015). Evaluation of the legal consequences of action affects neural activity and emotional experience during the resolution of moral dilemmas. *Brain and Cognition, 94*, 24–31.
- Prinz, J. J., & Nichols, S. (2010). Moral emotions. In J. M. Doris (Ed.), *The moral psychology handbook* (pp. 111–146). Oxford: Oxford University Press.
- Sarlo, M., Lotto, L., Manfrinati, A., Rumiati, R., Gallicchio, G., & Palomba, D. (2012). Temporal dynamics of cognitive-emotional interplay in moral decision-making. *Journal of Cognitive Neuroscience, 24*, 1018–1029.
- Sarlo, M., Lotto, L., Rumiati, R., & Palomba, D. (2014). If it makes you feel bad, don't do it! Egoistic rather than altruistic empathy modulates neural and behavioral responses in moral dilemmas. *Physiology & Behavior, 130*, 127–134.
- Sinnott-Armstrong, W. (1987). Moral realisms and moral dilemmas. *The Journal of Philosophy, 84*, 263–276.
- Szekely, R. D., & Miu, A. C. (2015). Incidental emotions in moral dilemmas: the influence of emotion regulation. *Cognition and Emotion, 29*, 64–75.
- Treadway, M. T., Buckholtz, J. W., Martin, J. W., Jan, K., Asplund, C. L., Ginter, M. R., Jones, O. D., & Marois, R. (2014). Corticolimbic gating of emotion-driven punishment. *Nature Neuroscience, 17*, 1270–1275.
- Wilson, T. D., & Gilbert, D. T. (2003). Affective forecasting. *Advances in Experimental Social Psychology, 35*, 345–411.