HARMING KIN TO SAVE STRANGERS: FURTHER EVIDENCE FOR ABNORMALLY UTILITARIAN MORAL JUDGMENTS AFTER VENTROMEDIAL PREFRONTAL DAMAGE

Bradley C. Thomas, **Katie E. Croft**, and **Daniel Tranel**
Department of Neurology, Division of Behavioral Neurology and Cognitive Neuroscience,
University of Iowa College of Medicine

Abstract

The ventromedial prefrontal cortex (vmPFC) has been implicated as a critical neural substrate mediating the influence of emotion on moral reasoning. It has been shown that the vmPFC is especially important for making moral judgments about "high-conflict" moral dilemmas involving direct personal actions, i.e., scenarios that pit compelling utilitarian considerations of aggregate welfare against the highly emotionally aversive act of directly causing harm to others (Koenigs, Young et al., 2007). The current study was designed to elucidate further the role of the vmPFC in high-conflict moral judgments, including those that involve *indirect* personal actions, such as indirectly causing harm to one's kin to save a group of strangers. We found that patients with vmPFC lesions were more likely than brain-damaged and healthy comparison participants to endorse utilitarian outcomes on high-conflict dilemmas regardless of whether the dilemmas (1) entailed direct versus indirect personal harms, and (2) were presented from the Self versus Other perspective. Additionally, all groups were more likely to endorse utilitarian outcomes in the Other perspective as compared to the Self perspective. These results provide important extensions of previous work, and the findings align with the proposal that the vmPFC is critical for reasoning about moral dilemmas in which anticipating the social-emotional consequences of an action (e.g., guilt or remorse) is crucial for normal moral judgments (Koenigs, Young et al., 2007; Greene 2007).

Keywords

SOCIAL COGNITION; EMOTION; PREFRONTAL CORTEX; MORAL

INTRODUCTION

Moral reasoning is one of the most exalted and important human mental faculties. The role of cognitive versus emotional processes in moral reasoning has been debated since Socrates. Kohlberg and Hume are the captains of the two sides in this debate, representing cognitive and emotional theories, respectively. As is the case with reasoning in general (Damasio, 1994), the importance of emotions in moral reasoning has gained strong empirical support in recent times. Converging research from functional imaging (Greene et al., 2001; Moll et al., 2002; Heekeren et al., 2003; Schaich Borg et al., 2006), behavioral (Wheatley & Haidt 2005; Valdesolo & DeSteno, 2006; Schnall et al., 2008), and lesion (Koenigs, Young et al., 2007; Ciaramelli et al., 2007) studies indicates a crucial role for emotions in moral cognition (for

Address for reply: Bradley C. Thomas, Department of Neurology, University of Iowa Hospitals and Clinics, 200 Hawkins Drive, Iowa City, Iowa 52242, bradley-thomas@uiowa.edu.

reviews, see Casebeer, 2003; Young & Koenigs, 2007). Neuroanatomically, the ventromedial prefrontal cortex (vmPFC) has been the focus of much of this work. The vmPFC is extensively connected to the amygdala, insula, hypothalamus, and brainstem, and has higher-order control over the execution of bodily components of emotions and the encoding of emotional values (Ongur & Price, 2000; Rolls, 2000). It has been suggested that the orbitofrontal cortex is involved in encoding the value of the consequences of actions (Rangel et al., 2008). Additionally, patients with vmPFC damage have impairments in social emotions and moral behavior in their everyday lives, for example, in the form of a lack of concern for others, socially inappropriate or callous behavior (e.g., inappropriate or rude comments), a reduction in the display of guilt and shame, and exhibiting increased aggressive behavior (Stuss & Benson, 1984; Eslinger & Damasio, 1985; Damasio et al., 1990; Rolls et al., 1994; Dimitrov et al., 1999; Anderson et al., 2006; Beer et al., 2006; Fellows et al., 2007).

The vmPFC has been implicated as a critical neural substrate mediating the influence of emotion on moral reasoning. It has been shown that the vmPFC is especially important for making moral judgments about "high-conflict" personal moral dilemmas that pit compelling considerations of aggregate welfare against highly emotionally aversive behaviors such as a direct, social-emotion arousing harm to others (Koenigs, Young et al., 2007). More generally, a moral dilemma creates a decision space that pits competing moral principles against each other. Drawing on the philosophical literature (Foot, 1978; Thompson, 1985), researchers have focused on utilitarian dilemmas that require choosing between a "utilitarian" outcome (which maximizes a certain variable, such as the number of lives saved) and a deontological outcome (which is consonant with a moral rule, such as "don't cause harm to others").

It is worth taking a philosophical aside at this point to explain the quotations around "utilitarian." There is a wide range of serious utilitarian philosophical theories, and the concept is used here only in a broad sense to capture the consequentialist, "life-maximization" aspect of the moral dilemmas being examined. This is merely to point out that very few utilitarians hold that the morally right thing to do is to simply maximize the number of lives saved and ignore all other variables (e.g., happiness, quality of life, potential for contribution to society; see Rosen (2003) for further discussion of the variety of classical utilitarian theories). We drop the quotations below, but maintain our sensitivity to this important point.

Previous research has examined different types of utilitarian dilemmas. One distinction that we will focus on involves the nature of the action required in order to secure the utilitarian outcome. To illustrate this, first consider the classic Footbridge dilemma, in which you are on a footbridge over some trolley tracks and see a run-away trolley heading toward five workers on the tracks (Thompson, 1985). The only way to save the workers is for you to push a large man off of the footbridge onto the tracks, where his weight will stop the train. Most people reject the choice of causing direct harm to this man, even though it maximizes the lives saved (Hauser et al., 2007). Now consider the Switch dilemma, which is like Footbridge except that you are standing next to the tracks and can flip a switch to turn the trolley onto a side-track—the five workers are saved, but the trolley then hits a single worker on the side-track (Foot, 1978). Here most people endorse causing the indirect harm by flipping the switch, sacrificing one life to save five (Hauser et al., 2007).

Greene and colleagues (2001) conducted a seminal neuroimaging study of moral judgments, and found increased vmPFC activity when participants made moral judgments of Footbridge-type dilemmas as compared to Switch-type dilemmas. In explaining why these dilemmas elicit different directions of utilitarian endorsement and differential vmPFC

activation, Greene et al. proposed "personal-ness," or the extent to which the proposed action arouses a social-emotional reaction, as a key factor—the direct physical harm in Footbridge is more "personal" than the indirect harm in Switch. Based on these findings and more recent work, Greene and colleagues have argued for a dual-process theory of moral judgment, which holds that moral judgments about personal actions are driven by social-emotional processing performed by the vmPFC, which prompts the rejection of direct personal harms (Greene 2007; Greene et al., 2004; Greene et al., 2008). In contrast, these researchers argue that judgments about impersonal harms are driven by the more cognitive operations of the dorsolateral prefrontal cortex, which prompts the endorsement of a utilitarian outcome.

In two neuropsychological studies, patients with vmPFC lesions (and severe defects in social emotions) performed similar to comparison participants on impersonal dilemmas, but on personal dilemmas they were significantly <u>more</u> likely to endorse utilitarian outcomes (e.g., pushing a man off of a bridge to save five lives; Koenigs, Young et al., 2007; Ciaramelli et al., 2007). However, Koenigs, Young, and colleagues also revealed a crucial nuance— patients with vmPFC lesions made more utilitarian judgments on personal dilemmas that were "high-conflict," like Footbridge, but not on dilemmas that were "low-conflict" but still personal (e.g., a dilemma that involves pushing your boss off of a building to get him out of your life). Both high- and low-conflict dilemmas were rated as having high emotional content; however, the high-conflict dilemmas involved conflict between aversion to direct, social-emotion arousing harm, on the one hand, and motivation to secure aggregate welfare, on the other. By contrast, the low-conflict dilemmas lacked this utilitarian motivation. In high-conflict dilemmas, the social-emotional properties (as prompted by a personal action such as having to kill one's baby, and the likely guilt that would normally be associated with that action), along with the fact that the action would maximize the number of lives saved, creates a high degree of competition between the choice options (Greene et al., 2004). As Greene (2007) summarized it, when a person considers the appropriateness of directly causing harm to secure a utilitarian outcome, as in a high-conflict dilemma, a vmPFC-mediated negative emotional response (or somatic-marker, cf. Damasio, 1994) suggests that the action is inappropriate. Since vmPFC patients lack this prepotent emotional response, they are more likely to judge the utilitarian action as appropriate (Koenigs, Young et al., 2007; Greene, 2007). More specifically, given the vmPFC's role in social emotions, and in anticipating the future consequences of an action and encoding the value of those consequences (Naqvi et al., 2006; Rangel et al., 2008), it seems likely that the relevant prepotent emotional responses (that are muted in vmPFC patients) involve anticipating the social-emotional consequences of the action, such as guilt for having performed it or shame as others express their moral disapproval. An abnormal judgment pattern results when this emotional information is not properly activated and factored into the decision, as seems to be the case for vmPFC patients, who have deficits in their ability to use emotion to guide decision making (e.g., Bechara et al., 1997).

However, the research discussed above leaves open a couple of interesting questions about the role of the vmPFC in judging high-conflict moral dilemmas, and our hypotheses for the current study build directly out of these unanswered questions. First, previous work has not examined high-conflict dilemmas in which the conflict is generated by an *indirectly-personal action*, in which a social-emotion arousing, personal harm is caused by an indirect rather than a direct action (Figure 1). Low-conflict dilemmas have been studied in which the action is directly-personal but lacking utilitarian justification, or in which the action is indirect and impersonal and thus lacking the social-emotional arousal elicited by personal actions. In both cases, one of the ingredients of a high-conflict dilemma is absent (either the utilitarian motivation or the personal-ness of the action). We reasoned that if indirect dilemmas involved personal, social-emotion arousing actions (such as killing one's daughter

or causing a gruesome, painful death), even without a direct, physical action, they would be high-conflict, so long as a utilitarian justification is also present. As Greene and colleagues (2001) noted, direct harms are not the only type of personal, social-emotion arousing actions. Thus, we sought to create high-conflict dilemmas that involved indirect harms.

The previously described version of the Switch dilemma exemplifies an indirect dilemma that is low-conflict. The utilitarian justification in that scenario is clear (one versus five deaths), and the harm required to secure the utilitarian outcome is indirect and impersonal (not strongly emotionally evocative), so the degree of conflict between the options is low and social-emotional information does not seem to be a critical factor. However, imagine a version of the Switch dilemma in which the one person on the side track is your daughter. Now, presumably, a high degree of conflict is introduced, as you are asked to choose between the death of your one daughter versus the deaths of five workers (all strangers). Such a dilemma involves an action that is indirect—but highly personal—along with a compelling utilitarian imperative, and we reasoned that social-emotional information likely becomes much more important for evaluating the options under this type of scenario (as compared to the low-conflict version of Switch as described above). Such indirectly-personal, high-conflict dilemmas have not been investigated previously. Based on the idea that the vmPFC is called into play for high-conflict dilemmas in general, and not just those that are direct, we hypothesized that *patients with vmPFC lesions would be significantly more likely than comparison participants to endorse utilitarian outcomes on high-conflict dilemmas of both the directly-personal and the indirectly-personal types*.

The second question we addressed relates to the perspective in which a moral dilemma is presented. In most previous work, dilemmas have been framed with the research participant as the proposed actor (we refer to these as "Self" dilemmas), and participants are asked to make judgments about their own hypothetical behavior (e.g., "Is it appropriate for *you* to push the man off the footbridge?"). The neural correlates of moral reasoning about dilemmas in which another person is described as the actor (which we will call "Other" dilemmas; e.g., "Is it appropriate for *John* to push the man off the footbridge?") have not been explored (Figure 1; see below). As we test the boundaries of vmPFC involvement in reasoning about moral dilemmas, exploring varied contexts like Self versus Other dilemmas is important. In this vein, if we are correct in reasoning that the vmPFC is critically involved in judging high-conflict dilemmas, then we would expect the vmPFC to be critical for making moral judgments about high-conflict dilemmas regardless of perspective (i.e., Self versus Other). Thus, we hypothesized that *patients with vmPFC damage would be significantly more likely than comparison participants to endorse utilitarian outcomes on both Self and Other high-conflict directly-personal dilemmas*.

Furthermore, a behavioral study revealed a Self-Other asymmetry in moral judgment about the Switch trolley dilemma (Nadelhoffer & Feltz, 2008). Specifically, participants were more likely to endorse flipping the switch in the Other condition than in the Self condition. Also, Berthoz et al. (2006) reported that a Self-Other contrast yielded greater amygdala activation for social/moral transgressions performed by the Self compared to Other, suggesting that the amygdala is especially important for emotional responses to one's own transgressions. Returning to our current study, we were interested in whether a Self-Other asymmetry would be observed for the high-conflict personal dilemmas—i.e., would participants be more likely to endorse a proposed aversive action when the actor was Other, versus Self? Predicting a Self-Other asymmetry is based on the idea that there would be greater aversion to imagining one's self causing a harm than imagining someone else causing a harm, akin to the finding that imagining one's self in pain is more emotionally arousing than imagining another person in pain (Jackson et al., 2006). And since this process does not seem to depend on the vmPFC (Berthoz et al., 2006), we hypothesized that *both*

comparison participants and participants with vmPFC lesions would demonstrate a Self-Other asymmetry, choosing more utilitarian options in the Other condition.

METHODS

Participants

A group of 9 patients with bilateral vmPFC damage (vmPFC group) and 9 brain-damaged comparison patients (with brain lesions located outside of the vmPFC and other emotion related areas, such as the amygdala and insula; BDC group) were recruited from the Patient Registry of the University of Iowa's Division of Behavioral Neurology and Cognitive Neuroscience. The vmPFC and BDC groups were comparable on demographic variables including age, education, and chronicity (time since lesion onset), and all patients were tested in the chronic epoch, 3 or more months post lesion onset (Table 1; although there was a 5.5 year difference in chronicity, this was not statistically significant, $t(16) = 1.60$, p = 0.13). All vmPFC (and no BDC) participants' neuropsychological profiles were remarkable for real-life social-emotional impairment (Table 2), in the presence of generally intact cognitive abilities (Table 3). A group of 11 neurologically normal comparison participants (NC group) was recruited from the local community (Table 1). One-way ANOVAs revealed no significant differences between the vmPFC, BDC, and NC groups on age ($F(2,26) = 0.006$, p = 0.994) or education ($F(2,26) = 0.622$, p = 0.545).

Neuroanatomical Analysis

The neuroanatomical analysis of vmPFC patients (Figure 2) was based on magnetic resonance data for six participants (those with lesions due to orbitofrontal meningioma resection) and on computerized tomography data for the other three participants (with lesions due to rupture of an anterior communicating artery aneurysm). All neuroimaging data were obtained in the chronic epoch. Each patient's lesion was reconstructed in three dimensions using Brainvox(Damasio & Frank, 1992; Frank et al., 1997). Using the MAP-3 technique described by Damasio et al. (2004), the lesion contour for each patient was manually warped into a normal template brain. The overlap of lesions in this volume, calculated by the sum of N lesions overlapping on any single voxel, is color-coded in Figure 2.

Experimental materials and methods

We crafted a battery of 29 novel moral dilemmas modeled on those previously used in the literature (Greene et al., 2001; Koenigs, Young et al., 2007). The dilemmas were presented in quasi-randomized order on a computer using the Presentation software package (Version 0.70, www.neurobs.com). They were divided into two blocks and the block order was counterbalanced between participants. For each dilemma, participants answered a moral and a comprehension question (as well as a believability and a punishment question not analyzed here). The *moral question* asked "How appropriate is it for [the agent] to [perform the proposed utilitarian action]?" For example, in the Self, indirectly-personal Switch scenario, the moral question asked, "How appropriate is it for you to flip the switch, killing your daughter to save the five workers?" Participants answered this question on a 6 point Likert scale from −3 to 3 with no zero response option (−3 = completely inappropriate; −1 = somewhat inappropriate; 1 = somewhat appropriate; 3 = completely appropriate). The *comprehension question* was designed to ensure that participants understood what was at stake in the dilemma and were able to hold that information online while making their judgment. This question asked a yes/no question about a specific feature of the dilemma, for example, "In the scenario you just read, were there 15 workers on the tracks?" Our study included stimuli that contain a variety of properties that may not be equally relevant to all participants (e.g., the scenario of killing one's own child, presumably more relevant to

persons with children than persons without children); however, we do not have reason to expect the relevance of these various properties to vary systematically across our participant groups, and the groups are so demographically similar that it is highly unlikely that such properties would have differential relevance to one group versus any other.

High-conflict direct versus indirect dilemmas

In the battery of dilemmas, we presented 5 high-conflict indirectly-personal moral dilemmas and 12 high-conflict directly-personal dilemmas, allowing for a comparison of indirect versus direct high-conflict dilemmas. High-conflict dilemmas are those that incorporate an emotionally charged, personal action (see *Emotion Ratings* below) as well as a utilitarian justification for performing the action (viz., maximizing the number of lives saved). All of these dilemmas were in the Self condition.

Self versus Other

In addition to the 12 high-conflict directly-personal *Self* dilemmas just mentioned, the dilemma battery also contained 12 high-conflict directly-personal *Other* dilemmas, which involved another person (e.g., "Bob") as the proposed actor. This manipulation allowed a comparison of Self and Other high-conflict dilemmas. Importantly, the Self and Other dilemmas had comparable hypothetical stakes; there was an equal number of Self and Other dilemmas requiring the killing of the participant's child in order to secure the utilitarian outcome; an equal number in which the participant's own life was at stake; and an equal number in which the participant's children's lives were at stake.

Emotion ratings

In order to determine whether the dilemmas differed in their emotional salience, we acquired emotion ratings from 8 of the normal comparison participants. (We only collected emotion ratings from normal participants, since these data were being used to ensure that the dilemmas were well equated on emotional salience.) For each of our 29 dilemmas, participants rated how emotionally upsetting the proposed social-emotional harm (e.g., pushing a person off of a footbridge) was on a 7 point Likert scale (1 = Not at all or slightly emotionally upsetting; 7 = Most emotionally abhorrent act I can imagine). All dilemmas had mean emotion ratings above 6, and a repeated measures ANOVA revealed no significant effect of Dilemma Type (indirect-self, direct-self, direct-other) on emotion ratings ($F(2,12) = 0.641$, $p = 0.544$). The dilemmas were well equated on their emotional salience.

Data Analysis

Participants' responses to the moral question were dichotomized for all analyses in order to reflect whether the participant endorsed the proposed utilitarian action. All negative responses, from −3 to −1, were coded as a rejection of the action, while all positive responses, from 1 to 3, were coded as an endorsement of the action.

Data from one NC participant were removed for missing 10 of the 29 comprehension questions. Of the remaining participants, one missed 7 comprehension questions and no others missed more than 4. For all analyses, we removed individual trials in which the participant missed the comprehension question. Also, two participants were removed for outlier response patterns. In both cases, the patterns were highly aberrant and unlike any we have seen in brain damaged patients or in healthy comparison participants: BDC 1290 was discarded for endorsing the utilitarian response in all directly-personal dilemmas, and one NC was discarded for endorsing the utilitarian response in all direct-other dilemmas while also having by far the highest proportion of endorsement for the other dilemma types.

We first determined whether there was a difference between the comparison groups' moral responses using a 3 (Dilemma Type: indirect-self, direct-self, direct-other) × 2 (Group: NC, BDC) repeated measures ANOVA with dichotomized responses to the moral question as the dependent variable. No main effect of Group ($F(1,15) = 1.474$, $p = 0.243$) or interaction effect ($F(1,15) = 0.36$, $p = 0.7$) was observed, and hence the NC and BDC groups were combined to form a single comparison (Comp) group to increase statistical power. We then compared the Comp and vmPFC groups using a similar 3 (Dilemma Type: indirect-self, direct-self, direct-other) × 2 (Group: vmPFC, Comp) repeated measures ANOVA. A main effect of Group and no interaction was predicted in accordance with hypotheses 1 and 2.

To further explore hypothesis 1, that the vmPFC group would be more likely than comparison participants to endorse both directly-personal and indirectly-personal utilitarian actions on high-conflict Self dilemmas, we ran a planned 2 (Dilemma Type: indirect-self, direct-self) × 2 (Group: vmPFC, Comp) repeated measures ANOVA. A main effect of Group was predicted, but not a main effect of dilemma type nor an interaction.

To further explore hypotheses 2 and 3, that the vmPFC group would be more likely to endorse utilitarian outcomes on both Self and Other dilemmas and that a Self-Other bias would be observed in all participants, we ran a planned 2 (Perspective: direct-self, direct-other) × 2 (Group: vmPFC, Comp) repeated measures ANOVA, with main effects of Group (hypothesis 2) and Perspective (hypothesis 3), but no interaction, predicted.

RESULTS

Patients with vmPFC damage show increased utilitarian judgments

In support of hypotheses 1 and 2, the 3 (Dilemma Type: indirect-self, direct-self, direct-other) × 2 (Group: vmPFC, Comp) repeated measures ANOVA yielded a main effect of Group ($F(1,24) = 6.639$, $p = 0.017$), and no effect of Dilemma Type nor an interaction, revealing that the vmPFC group was more likely than the Comp group to endorse utilitarian outcomes (see Figure 3).

High-conflict direct versus indirect dilemmas

In support of hypothesis 1, the planned 2 (Dilemma Type: indirect-self, direct-self) × 2 (Group: vmPFC, Comp) repeated measures ANOVA yielded a main effect of Group ($F(1,24) = 5.67$, $p = 0.026$; Figure 3), reflecting the fact that the vmPFC group endorsed utilitarian actions more frequently than the comparison group. There was a marginally significant main effect of Dilemma Type ($F(1,24) = 3.11$, $p = 0.09$), and no interaction effect ($p = 0.668$), indicating that the vmPFC > Comp effect is not affected by whether the high-conflict dilemmas are indirect or direct.

Self versus Other

In support of hypothesis 2, the planned 2 (Perspective: direct-self, direct-other) × 2 (Group: vmPFC, Comp) repeated measures ANOVA yielded a main effect of Group, revealing that the vmPFC group was more likely than the Comp group to endorse utilitarian outcomes ($F(1,24) = 5.427$, $p = 0.029$). And in support of hypothesis 3, there was a significant effect of Perspective, with both groups being more likely to endorse the utilitarian outcome in Other as compared to Self dilemmas, revealing a Self-Other bias ($F(1,24) = 17.455$, $p < 0.001$; Figure 3). Lastly, the interaction effect was not significant ($p = 0.966$), suggesting that the Self-Other bias was of the same magnitude for both the Comp and vmPFC groups.

DISCUSSION

Using a novel set of moral dilemmas, we replicated the finding that the vmPFC is critical for making normal moral judgments about high-conflict directly-personal dilemmas (Koenigs, Young et al., 2007). Importantly, we have also extended this to indirectly-personal dilemmas, and these new findings indicate that the vmPFC is a critical neural substrate for making normal moral judgments about both direct and indirect high-conflict dilemmas. Moreover, we have extended this finding beyond first-person dilemmas to show that the vmPFC is critical for normal moral judgments about both Self and Other directly-personal dilemmas. Thus, it seems that the vmPFC is critical for normal reasoning about moral dilemmas when there is a high degree of conflict and social emotions are important factors in evaluating the choice-options, regardless of whether the dilemmas are direct versus indirect, or Self versus Other. This is an important extension of previous work by Greene and colleagues (2004) and from our laboratory (Koenigs, Young, et al., 2007), and it puts further empirical teeth in the ideas set forth by Koenigs, Young et al. (2007) and by Greene (2007), in regard to the neural mechanisms behind these phenomena.

Greene et al. (2001) initially suggested that increased emotional salience drove vmPFC involvement in moral judgments about directly-personal dilemmas. Subsequent lesion work bolstered and refined this idea, suggesting that the crucial factor (vis-à-vis vmPFC involvement) was that the dilemmas were high-conflict—i.e., there was a high degree of competition between the choice options due to conflict between a personal, social-emotional harm and a utilitarian outcome. This conflict creates a demand to integrate social-emotional information into the decision-making process (Koenigs, Young et al., 2007; Greene 2007), thus prompting vmPFC involvement. In the present study, we have provided further support for this notion by showing that the vmPFC is critical in the formation of normal moral judgments about such high-conflict moral dilemmas, regardless of whether the social-emotion arousing harmful action is of the direct or indirect type.

Participants were marginally more likely to endorse utilitarian outcomes in indirectly- as compared to directly-personal dilemmas, suggesting that the direct nature of the proposed actions may have influenced participants' moral judgments (despite our having attempted to equate the personal-ness and emotional salience of the direct and indirect dilemmas). Greene and colleagues have recently argued that moral judgment is strongly influenced by whether the action being judged involves "personal force" and intention, which is a property of most of our direct moral dilemmas (Greene et al., 2009), and this may account for the marginal direct-indirect difference in moral judgment—the personal force involved in most direct actions may have driven participants away from the utilitarian response more than the indirect actions which lacked personal force. Considering the non-significant interaction effect in the present experiment (vmPFC patients were swayed by the indirect-direct distinction to a comparable degree as comparison participants), it seems that the vmPFC may not be crucial for the effect of "physical actions" on moral judgment.

Moral reasoning has often been viewed as a special faculty, marked by its dependence on accounting for the interests of others, in contrast to other reasoning processes that focus on self-interest, such as economic or ordinary reasoning. Moral reasoning does involve processing social information, e.g., about others' mental states and intentions (Casebeer, 2003; Young et al., 2007; Young et al., 2010). However, there are important features that moral reasoning has in common with other types of reasoning (Reyna & Casillas, 2009). In particular, the vmPFC's role in judging high-conflict moral dilemmas seems to be essentially the same role it plays in other complex reasoning processes. For example, in complex decision making with pecuniary implications, such as on the Iowa Gambling Task, the vmPFC is critical for integrating emotional information into the reasoning process

(Bechara et al., 1997; Naqvi et al., 2006). In particular, the vmPFC plays this role when the anticipation of future consequences is used to form a judgment (Naqvi et al., 2006). The high-conflict moral dilemmas studied here, both direct and indirect, have this feature as their hallmark; the anticipation of serious social-emotional consequences of the proposed personal action (e.g., living with the knowledge that you killed your daughter or pushed a person off of a footbridge) seems to be required to reason normally about these dilemmas. By anticipation we do not mean to suggest that the emotion is not presently experienced, but only that the emotion is a projection into the future based on how one would feel if she performed the action rather than an emotion that is based on a decision already made or an action already performed. Although this "anticipation hypothesis" is our preferred interpretation, our data do not directly adjudicate between this interpretation and one in which the emotion is not anticipatory (e.g., a feeling of disgust elicited by reading about the aversive personal action). Either way, in contrast to high-conflict dilemmas, there is a clear lack of such social-emotional demands (anticipatory or not) in low-conflict dilemmas. Although directly-personal low-conflict dilemmas (e.g., Architect described in Figure 1) involve social-emotional harms, this information is not critical to the decision process as there is no countervailing force or competition in the choice options—one need not be driven by the social-emotional information in order to respond normally in such cases. Also in support of the notion that the vmPFC's involvement in moral judgment is similar to its role in other forms of decision making, behavioral studies have shown moral judgments to be susceptible to many of the same biases as other judgment and decision making processes (Tversky & Kahneman, 1981; Petrinovich & O'Neill, 1996; Uhlmann et al., 2009). The Self-Other asymmetry revealed by the present study is another example of a bias that affects many other types of judgment and also influences moral judgment (Jones & Nisbitt, 1972; Nadelhoffer & Feltz, 2008).

The suggestion that moral reasoning involves some of the same mechanisms as other complex forms of reasoning is by no means new (Reyna & Casillas, 2009), and it does not diminish the important and special role moral reasoning plays in the human mind and in human societies. It does, however, suggest that, at least at a gross neural level, moral reasoning may not be so different from other complex reasoning processes. More research is warranted to further explore this issue. It is critical to understand the psychological processes and neural mechanisms involved in moral reasoning and how they may differ from other reasoning processes, and how they may vary as moral reasoning occurs in varying contexts. For example, it is unclear whether the involvement of the vmPFC in high-conflict utilitarian moral dilemmas will generalize to other types of moral dilemmas and to other, non-dilemmatic moral contexts. In regards to this point, one specific limitation of the present study is that we did not compare Self and Other indirectly-personal dilemmas. Future research should test such dilemmas to determine if our observed pattern of results holds up, suggesting, as we would predict, that the vmPFC is critical for reasoning about Self and Other high-conflict dilemmas of the indirect type in addition to the direct type. An additional limitation of the present study is the use of hypothetical moral dilemmas. Although such dilemmas are commonly used in moral reasoning research, they are far removed from real-world moral scenarios, and more work is needed on the neural basis of real-world and everyday moral reasoning (Monin et al., 2007). The lesion method will play an important role in this line of research because it is well suited for studying complex human behavior in real-world contexts.

Finally, extending the work of Nadelhoffer and Feltz (2008) and Berthoz et al. (2006), our findings reveal a Self-Other asymmetry in moral judgments about high-conflict dilemmas. And we further show that this asymmetry is insensitive to vmPFC damage, as all groups showed the same magnitude of Self-Other asymmetry, endorsing significantly more utilitarian actions in response to dilemmas in the Other condition as compared to those in the

Self condition. In other words, participants were more likely to judge it appropriate for "Bob" than for the participant herself to cause the harm in order to secure the utilitarian outcome. Thus, shifting from judging the appropriateness of one's own actions to judging the actions of others results in the adoption of a more utilitarian stance, a phenomenon similar to what utilitarian philosophers call "impersonal benevolence," in which utilitarian judgments result when one makes moral judgments without regard for her own personal stakes (for a philosophical discussion of such Self-Other asymmetries in ethics see Slote, 1984). The Self-Other asymmetry in moral judgment may be due to a simple increased aversion to imagining one's self causing the harm as compared to imagining another person causing the harm. This idea is compatible with functional neuroimaging findings from Berthoz et al. (2006), who emphasized the amygdala as having a key role in the Self-Other distinction in moral cognition—the amygdala, more so than the vmPFC, is strongly implicated in basic emotions and arousal of the type that may be fundamental to the Self-Other asymmetry we have found. One limitation of the present study that warrants caution in regards to the observed Self-Other asymmetry is that we did not use identical dilemmas in the Self and Other conditions, meaning that it is possible that a difference in the dilemmas between conditions is responsible for the asymmetry in judgments. However, this does not seem to be the best explanation of the finding, as the Self and Other dilemmas were equated on several critical factors, like the emotional salience of the proposed actions, how often they involved risk to the lives of the participants and/or their children, and how often they required the death of the participants' own children in order to secure the utilitarian outcome.

In sum, we show that patients with vmPFC lesions are more likely than comparison participants to endorse utilitarian outcomes on high-conflict dilemmas (1) regardless of whether the dilemmas are of the direct or indirect type, and (2) regardless of whether directly-personal dilemmas are in the Self or Other perspective. Moreover, both comparison participants and participants with vmPFC lesions were more likely to endorse utilitarian outcomes on high-conflict directly-personal dilemmas framed in the Other perspective as compared to those framed in the Self perspective. Taken together, these results suggest that the vmPFC is critical for reasoning about moral dilemmas in which anticipating the social-emotional consequences of an action is important for forming a moral judgment.

Acknowledgments

References

Anderson S, Barrash J, Bechara A, Tranel D. Impairments of emotion and real-world complex behavior following childhood- or adult-onset damage to ventromedial prefrontal cortex. Journal of the International Neuropsychological Society. 2006; 12:224–235. [PubMed: 16573856]

Barrash J, Tranel D, Anderson S. Acquired personality disturbances associated with bilateral damage to the ventromedial prefrontal region. Developmental Neuropsychology. 2000; 18:355–381. [PubMed: 11385830]

Bechara A, Damasio H, Tranel D, Damasio AR. Deciding advantageously before knowing the advantageous strategy. Science. 1997; 275:1293–1295. [PubMed: 9036851]

Beer J, Oliver JP, Scabini D, Knight RT. Orbitofrontal cortex and Social behavior: Integrating self-monitoring and emotion–cognition interactions. Journal of Cognitive Neuroscience. 2006; 18:871–879. [PubMed: 16839295]

Berthoz A, Grezes J, Armony JL, Passingham RE, Dolan RJ. Affective response to one's own moral violations. NeuroImage. 2006; 31:945–950. [PubMed: 16490367]

Casebeer WD. Moral cognition and its neural constituents. Nature Reviews Neuroscience. 2003; 4:840–846.

Ciaramelli E, Muccioli M, Làdavas E, di Pellegrino G. Selective deficit in personal moral judgment following damage to ventromedial prefrontal cortex. SCAN. 2007; 2:84–92. [PubMed: 18985127]

Damasio, AR. Descartes' error: Emotion, rationality and the human brain. New York: Putnam; 1994.

Damasio AR, Tranel D, Damasio H. Individuals with sociopathic behavior caused by frontal damage fail to respond autonomically to social stimuli. Behavioral Brain Research. 1990; 41:81–94.

Damasio H, Frank R. Three-dimensional in vivo mapping of brain lesions in humans. Archives of Neurology. 1992; 49:137–143. [PubMed: 1736845]

Damasio H, Tranel D, Grabowski T, Adolphs R, Damasio AR. Neural systems behind word and concept retrieval. Cognition. 2004; 92:179–229. [PubMed: 15037130]

Dimitrov M, Phipps M, Zahn TP, Grafman J. A thoroughly modern gage. Neurocase. 1999; 5:345–354.

Eslinger PJ, Damasio AR. Severe disturbance of higher cognition after bilateral frontal lobe ablation: Patient EVR. Neurology. 1985; 35:1731–1741. [PubMed: 4069365]

Fellows LK. Advances in understanding ventromedial prefrontal function: The accountant joins the executive. Neurology. 2007; 68:991–995. [PubMed: 17389302]

Frank RJ, Damasio H, Grabowski TJ. Brainvox: an interactive, multimodal visualization and analysis system for neuroanatomical imaging. Neuro Image. 1997; 5:13–30. [PubMed: 9038281]

Foot, P. Virtues and vices and other essays in moral philosophy. Berkeley: University of California Press; 1978. The Problem of abortion and the doctrine of double effect; p. 19-32.(Reprinted from *Oxford Review 5*, 5–15, 1967.)

Greene JD. Why are VMPFC patients more utilitarian?: A dual-process theory of moral judgment explains. Trends in Cognitive Sciences. 2007; 11:322–323. [PubMed: 17625951]

Greene JD, Cushman FA, Stewart LE, Lowenberg K, Nystrom LE, Cohen JD. Pushing moral buttons: The interaction between personal force and intention in moral judgment. Cognition. 2009; 111:364–371. [PubMed: 19375075]

Greene JD, Morelli SA, Lowenberg K, Nystrom LE, Cohen JD. Cognitive load selectively interferes with utilitarian moral judgment. Cognition. 2008; 107:1144–1154. [PubMed: 18158145]

Greene JD, Nystrom LE, Engell AD, Darley JM, Cohen JD. The neural bases of cognitive conflict and control in moral judgment. Neuron. 2004; 44:389–400. [PubMed: 15473975]

Greene JD, Sommerville RB, Nystrom LE, Darley JM, Cohen JD. An fMRI investigation of emotional engagement in moral judgment. Science. 2001; 293:2105–2108. [PubMed: 11557895]

Hauser M, Cushman F, Young L, Kang-Xing Jin R, Mikhail J. A dissociation between moral judgments and justifications. Mind & Language. 2007; 22:1–21.

Heekeren HR, Wartenburger I, Schmidt H, Schwintowski HP, Villringer A. An fMRI study of simple ethical decision-making. NeuroReport. 2003; 14:1215–1219. [PubMed: 12824762]

Jackson PL, Brunet E, Meltzoff AN, Decety J. Empathy examined through the neural mechanisms involved in imagining how I feel versus how you would feel pain: An event-related fMRI study. Neuropsychologia. 2006; 44:752–761. [PubMed: 16140345]

Jones, EE.; Nisbett, RE. The actor and the observer: Divergent perceptions of the causes of behavior. In: Jones, EE.; Kanouse, D.; Kelley, HH.; Nisbett, RE.; Valins, S.; Weiner, B., editors. Attribution: Perceiving the causes of behavior. Morristown: General Learning Press; 1972. p. 79-94.

Koenigs M, Young L, Adolphs R, Tranel D, Cushman F, Hauser M, Damasio AR. Damage to the prefrontal cortex increases utilitarian moral judgments. Nature. 2007; 446:908–911. [PubMed: 17377536]

Moll J, de Oliveira-Souza R, Bramati IE, Grafman J. Functional networks in emotional moral and nonmoral social judgments. NeuroImage. 2002; 16:696–703. [PubMed: 12169253]

Monin B, Pizarro DA, Beer JS. Deciding versus reacting: Conceptions of moral judgment and the reason-affect debate. Review of General Psychology. 2007; 11:99–111.

Nadelhoffer T, Feltz A. The actor–observer bias and moral intuitions: Adding fuel to Sinnott-Armstrong's fire. Neuroethics. 2008; 1:133–144.

Naqvi N, Shiv B, Bechara A. The role of emotion in decision making: A cognitive neuroscience perspective. Current Directions in Psychological Science. 2006; 15:260–264.

Ongur D, Price J. The organization of networks within the orbital and medial prefrontal cortex of rats, monkeys, and humans. Cerebral Cortex. 2000; 10:206–219. [PubMed: 10731217]

Petrinovich L, O'Neill P. Influence of wording and framing effects on moral intuitions. Ethology and Sociobiology. 1996; 17:145–171.

Rangel A, Camerer C, Montague PR. A framework for studying the neurobiology of value-based decision making. Nature Reviews Neuroscience. 2008; 9:545–556.

Reyna, VF.; Casillas, W. Development and dual processes in moral reasoning: A fuzzy-trace theory approach. In: Bartels, DM.; Bauman, CW.; Skitka, LJ.; Medin, DL., editors. Moral judgment and decision making: The psychology of learning and motivation. Vol. 50. San Diego: Elsevier Academic Press; 2009. p. 207-236.

Rolls ET. The orbitofrontal cortex and reward. Cerebral Cortex. 2000; 3:284–294. [PubMed: 10731223]

Rolls ET, Hornak J, Wade D, McGrath J. Emotion-related learning in patients with social and emotional changes associated with frontal lobe damage. Journal of Neurology, Neurosurgery & Psychiatry. 1994; 57:1518–1524.

Rosen, F. Classical utilitarianism from Hume to Mill. London: Routledge; 2003.

Schaich Borg J, Hynes C, Van Horn J, Grafton S. Consequences, action, and intention as factors in moral judgments: An fMRI investigation. Journal of Cognitive Neuroscience. 2006; 18:803–817. [PubMed: 16768379]

Schnall S, Haidt J, Clore GL, Jordan AH. Disgust as embodied moral judgment. Personality and Social Psychology Buletin. 2008; 34:1096–1109.

Slote M. Morality and self-other asymmetry. The Journal of Philosophy. 1984; 81:179–192.

Stuss DT, Benson DF. Neuropsychological studies of the frontal lobes. Psychological Bulletin. 1984; 95:3–28. [PubMed: 6544432]

Thomson JJ. The trolley problem. Yale Law Journal. 1985; 94:1395–1415.

Tversky A, Kahneman D. The framing of decisions and the psychology of choice. Science. 1981; 211:453–458. [PubMed: 7455683]

Uhlmann EL, Pizarro DA, Tannenbaum D, Ditto PH. The motivated use of moral principles. Judgment and Decision Making. 2009; 4:476–491.

Valdesolo P, DeSteno D. Manipulations of emotional context shape moral judgment. Psychological Science. 2006; 17:476–477. [PubMed: 16771796]

Wheatley T, Haidt J. Hypnotic disgust makes moral judgments more severe. Psychological Science. 2006; 16:780–784. [PubMed: 16181440]

Young L, Bechara A, Tranel D, Damasio H, Hauser M, Damasio A. Damage to ventromedial prefrontal cortex impairs judgment of harmful intent. Neuron. 2010; 65:845–851. [PubMed: 20346759]

Young L, Cushman F, Hauser M, Saxe R. The neural basis of the interaction between theory of mind and moral judgment. Proceedings of the National Academy of Sciences. 2007; 104:8235–8240.

Young L, Koenigs M. Investigating emotion in moral cognition: a review of evidence from functional neuroimaging and neuropsychology. British Medical Bulletin. 2007; 84:69–79. [PubMed: 18029385]
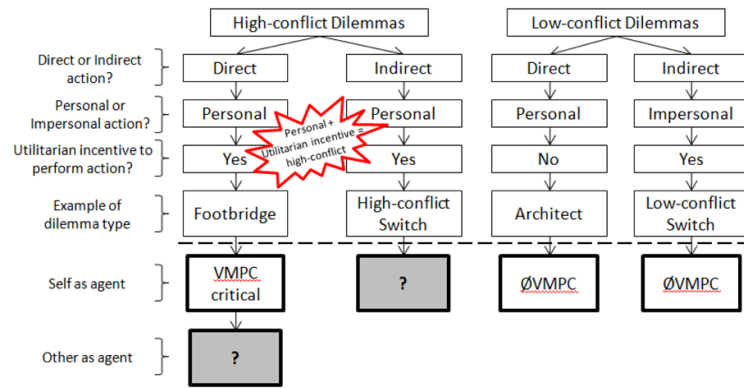
Figure 1.
Typology of moral dilemmas relevant to the present study. Grey boxes represent the contributions of the current study. A summary of results from previous studies on vmPFC involvement in moral judgments about each dilemma type is shown below the dashed line. Low-conflict dilemmas have been examined in previous studies and are not explored in the present work. High-conflict dilemmas involve personal actions and utilitarian justifications, pitting an action that results in a highly aversive, social-emotion arousing harm (i.e., a personal action, such as a directly or indirectly caused aversive harm) against a utilitarian outcome. High-conflict directly-personal dilemmas in the Self condition have been examined in previous studies; here, we added the Other condition. High-conflict indirectly-personal dilemmas are new in the present study. (Two additional types of low-conflict dilemmas are theoretically possible, and we predict would not place demands on the vmPFC, but they were not considered in the present study: (1) a direct, impersonal dilemma (in which the direct harm does not evoke a strong social-emotional reaction) with a utilitarian justification (see Greene et al., 2009 for possible examples) and (2) an indirect, impersonal dilemma without a utilitarian justification.) Example dilemmas: "Footbridge" involves directly pushing a man to his death to save 5 workers. "Switch" involves flipping a switch, indirectly killing your daughter (high-conflict) or a stranger (low-conflict), to save 5 workers. "Architect" involves pushing your boss off of a building to get him out of your life (Greene et al., 2001).

Right Hemisphere Mesial View   Frontal View   Left Hemisphere Mesial View
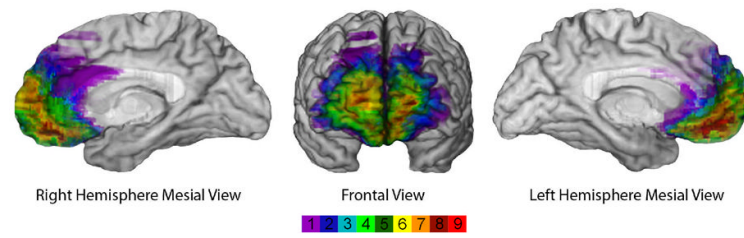
1 2 3 4 5 6 7 8 9

Figure 2.
Lesion overlap of vmPFC patients. Mesial and frontal views of the overlap map of lesions for the 9 vmPFC patients. The color bar indicates the number of overlapping lesions at each voxel. The area of maximal overlap lies in the ventromedial prefrontal cortex.
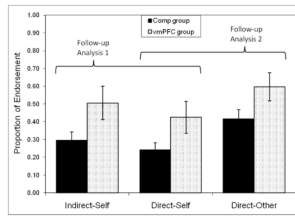
Figure 3.
Moral judgments for high-conflict dilemmas of the indirect-self, direct-self, and direct-other type. The proportion of endorsement of the utilitarian option is shown, with error bars reflecting standard error of the mean. The proportion of endorsement was significantly greater for the vmPFC group than the comparison group across all three types of dilemmas, and no interaction effect was observed. In planned follow-up analyses, we found that: (1) When comparing indirect-self versus direct-self dilemmas, there was a main effect of group (vmPFC > Comp) and a trend toward higher proportion of endorsement on indirect as compared to direct dilemmas, but no interaction. (2) When comparing direct-self versus direct-other dilemmas, there was a main effect of group (vmPFC > Comp) and a main effect of perspective (Other > Self), but no interaction.

Table 1

Demographic and clinical data. Individual participants are in the vmPFC group. Group means and standard deviations (SD) are reported below. Age is in years at time of testing. Edu. is education in years of formal schooling. Chronicity is the time between lesion onset and completion of the present experiment, in years. Handedness reports dominant hand. Etiology describes the cause of neurological lesion (SAH = subarachnoid hemorrhage; ACoA = anterior communicating artery). BDC patients had brain damage due to surgical intervention (n = 4) or cerebrovascular disease (n = 5).

Patient	Gender	Age	Edu.	Hand.	Chron.	Etiology
318	M	68	14	R	32	Meningioma resection
770	F	66	16	R	23	Meningioma resection
1983	F	45	14	R	12	SAH
2352	F	59	14	R	9	SAH; ACoA aneurysm
2391	F	62	12	R	7	Meningioma resection
2577	M	68	11	R	9	SAH; ACoA aneurysm
3032	M	52	12	R	6	Meningioma resection
3349	F	66	12	L	3	Meningioma resection
3350	M	56	18	R	3	Meningioma resection
vmPFC						
Mean (SD)	4M;5F	60.2 (8.0)	13.7 (2.2)	8R;1L	11.6 (9.75)	
BDC						
Mean (SD)	5M;4F	60.2 (11.2)	14.6 (2.7)	8R;1L	6.0 (3.7)	
NC						
Mean (SD)	6M;5F	59.8 (8.5)	14.9 (2.6)	9R;1L; 1 Mixed		

Table 2

Emotional and social functioning data for vmPFC patients. SCR denotes skin conductance responses to emotionally charged stimuli (e.g., pictures of social disasters, mutilations, and nudes, using methods described previously; Damasio et al., 1990). None of the 9 BDC patients had SCR impairments to emotionally charged stimuli. Social Emotions, the patient's demonstrated capacity for empathy, embarrassment, and guilt, as determined from reports from a collateral source (spouse or family member) provided on the Iowa Scales of Personality Change (Barrash et al., 2000) and from data from clinical interviews. Acquired Personality Changes, post-lesion changes in personality (e.g., irritability, emotional dysregulation, and impulsivity), as determined from data from the Iowa Scales of Personality Change. For Social Emotions and Acquired Personality Changes, the degree of severity is designated in parentheses (1 = mild, 2 = moderate, 3 = severe).

Patient	SCRs	Social Emotions	Acquired Personality Changes
318	Impaired	Defective (3)	Yes (3)
770	Impaired	Defective (3)	Yes (3)
1983	Impaired	Defective (3)	Yes (3)
2352	Impaired	Defective (2)	Yes (3)
2391	Impaired	Defective (3)	Yes (2)
2577	Impaired	Defective (3)	Yes (3)
3032	Impaired	Defective (3)	Yes (2)
3349	n/a	Defective (1)	Yes (1)
3350	Impaired	Defective (3)	Yes (3)

Table 3

Neuropsychological data for vmPFC patients. WAIS-III, Wechsler Adult Intelligence Scale-III score (VIQ, verbal IQ; PIQ, performance IQ; FSIQ, full-scale IQ; WMI, working memory index; for all four indices, 90–109 is average, 110–119 is high average, 120+ is superior). AVLT Recall, z-score on the Auditory Verbal Learning Test 30-minute delayed recall (scores 1.5 standard deviations below normal, or scores of −1.5 or below, are impaired). BDAE, Boston Diagnostic Aphasia Examination (CIM, Complex Ideational Material subtest, a measure of auditory comprehension, number correct out of 12; RSP, Reading Sentences and Paragraphs subtest, a measure of reading comprehension, number correct out of 10). TMT B − A, z-score of the Trail-Making Test Part B minus Part A times to completion (scores −1.5 or below are impaired). MMPI-2 D, T score, Minnesota Multiphasic Personality Inventory-2 Depression scale (T-scores 1.5 standard deviations above the mean, 65 and higher, are impaired).

| Patient | WAIS-III | | | | AVLT | BDAE | | TMT | MMPI-2 D |
	VIQ	PIQ	FSIQ	WMI	Recall	CIM	RSP	B-A	
318	142	134	143	119	1	12	10	0.9	51
770	119	94	108	113	3	12	9	−0.5	73
1983	110	105	108	99	−1.6	n/a	10	1.6	40
2352	108	102	106	111	1.4	n/a	10	1.6	42
2391	110	107	109	104	2.6	n/a	9	1.4	55
2577	89	80	84	80	0.3	12	6	−1.8	64
3032	94	113	102	95	−2.6	12	10	−1	70
3349	101	100	101	104	0.3	12	9	−1.6	n/a
3350	119	113	118	121	−4.2	12	10	−0.8	52