Selective deficit in personal moral judgment following damage to ventromedial prefrontal cortex

Elisa Ciaramelli,[1,2] Michela Muccioli,[2] Elisabetta Làdavas,[1,2] and Giuseppe di Pellegrino[1,2]

[1]Dipartimento di Psicologia, Università di Bologna, Bologna, Italy, and [2]Centro Studi e Ricerche di Neuroscienze Cognitive, Cesena, Italy

Recent fMRI evidence has detected increased medial prefrontal activation during contemplation of personal moral dilemmas compared to impersonal ones, which suggests that this cortical region plays a role in personal moral judgment. However, functional imaging results cannot definitively establish that a brain area is necessary for a particular cognitive process. This requires evidence from lesion techniques, such as studies of human patients with focal brain damage. Here, we tested 7 patients with lesions in the ventromedial prefrontal cortex and 12 healthy individuals in personal moral dilemmas, impersonal moral dilemmas and non-moral dilemmas. Compared to normal controls, patients were more willing to judge personal moral violations as acceptable behaviors in personal moral dilemmas, and they did so more quickly. In contrast, their performance in impersonal and non-moral dilemmas was comparable to that of controls. These results indicate that the ventromedial prefrontal cortex is necessary to oppose personal moral violations, possibly by mediating anticipatory, self-focused, emotional reactions that may exert strong influence on moral choice and behavior.

INTRODUCTION

Recent interest in social cognitive neuroscience has led to a growing body of research aimed at elucidating the neural and cognitive mechanisms that underlie human moral behavior (see for recent reviews Moll *et al.*, 2005; Beer and Ochsner, 2006; Lieberman, 2006; McKinnon *et al.*, 2006). Moral behavior refers to what individuals should do based on principles and judgments (i.e. moral values) shared with other members of their social environment. The initial foray into the neuroscience of moral behavior and reasoning came from the systematic examinations of changes in the social life of people with localized brain damage produced by accidents, strokes or neurological disease. Patients with lesions in the orbitofrontal and ventromedial prefrontal cortex have long been described as presenting high levels of aggressiveness, lack of concern for social and moral rules and irresponsibility (e.g. Eslinger and Damasio, 1985; Stuss *et al.*, 1992; Damasio, 1994; Blair and Cipolotti, 2000), which suggests these brain areas are important neural correlates of moral behavior. Lesions of the same areas during childhood impair the development of moral knowledge and ethical judgment (Anderson *et al.*, 1999), further suggesting that these brain regions are important neural correlates of moral behavior. Accordingly, functional imaging studies in healthy individuals involving moral judgment (Moll *et al.*, 2002a; Heekeren *et al.*, 2003) and moral reasoning

(Greene *et al.*, 2001; 2004; Borg *et al.*, 2006) have detected consistent activations of the orbitofrontal and ventromedial prefrontal cortex, although activated regions encompass a more extended network of neural regions (see Moll *et al.*, 2005 and references therein).

One crucial question concerns the specific mechanism by which these frontal regions promote behaviors that conform to, rather than violate, moral values and expectations shared by a social group. According to classical moral theories, moral behavior is a perfectly rational type of affair, governed by deliberative and high cognitive processing. A more recent view, however, emphasizes the role of intuitive and affective processes in social and ethical decision-making (Damasio, 1994; Greene *et al.*, 2001).

Consistent with this latter view, Greene and colleagues have proposed that medial prefrontal areas might mediate strong negative emotional responses to moral violations, which prevent individuals from implementing such morally impermissible actions (Greene and Haidt, 2002). These emotions might be the by-product of (or, alternatively, evolved to promote) humans' intensely social nature, which relies on behaviors warranting the cohesion of social groups (Greene, 2003). In a series of fMRI experiments (Greene *et al.*, 2001; 2004), the authors explored this possibility by studying healthy individuals who were considering moral dilemmas. Ethicists have called moral dilemmas situations in which a person faces a conflict between two (or more) opposing moral values or requirements. Specifically, Greene *et al.* (2001) compared individuals' performance on two different types of moral dilemmas, i.e. those involving 'personal' and those involving 'impersonal'

moral judgments. A typical personal moral dilemma involves having to decide whether or not to push a stranger off of a footbridge in front of an oncoming trolley in order to save five people on the main track (i.e. the *footbridge dilemma*). In a quite similar situation, an impersonal moral dilemma involves having to decide whether or not to hit a switch that will turn the trolley to an alternate set of tracks, where it will kill one person instead of five (i.e. the *trolley dilemma*). In both types of dilemmas individuals are required to judge whether it is appropriate to incur in a moral violation (i.e. killing one person) in order to maximize overall consequences (i.e. saving five persons). However, whereas personal moral violations consist in (i) causing serious bodily harm (ii) to a human being (iii) through one's own agency (i.e. in such a way that the harm does not result from the deflection of an existing threat onto a different party), impersonal moral violations do not satisfy at least one of these criteria (e.g. #3 in the case of the *trolley dilemma*), and, therefore, may induce a less intense emotional experience in individuals (Greene *et al.*, 2001).

Greene and colleagues found that medial prefrontal regions commonly associated with social/emotional processing (Damasio, 1994; Berthoz *et al.*, 2002; Moll *et al.*, 2002b; 2005), including the medial prefrontal gyrus and the posterior cingulate gyrus, were strongly activated while responding to personal, but not impersonal, moral dilemmas (Greene *et al.*, 2001). Importantly, this medial prefrontal activation appeared to interfere with 'utilitarian' moral judgment: Individuals were slower to approve, compared to refuse, personal moral violations, consistent with the idea that the approval of a personal moral violation is in conflict with emotional intuitions, whereas its refusal is a rather automatic reaction (Greene and Haidt, 2002). This pattern of results was not detected for impersonal moral judgments or dilemmas with no moral connotation (non-moral dilemmas), suggesting these were mainly accomplished through logical reasoning, supported by dorsolateral prefrontal cortex (Greene *et al.*, 2004), with relatively scarce contribution from processes dedicated to social cognition.

Although these neuroimaging studies have suggested a role of the medial prefrontal cortex in personal moral judgment, it is currently unclear whether this brain region is *essential* for determining normal moral behavior, or is co-activated with the crucial region, but contributes little, if anything. In this respect, a stronger case could be made if one uncovered patients with medial prefrontal lesions who show abnormal personal moral judgment. Thus, in order to integrate lesion data with the neuroimaging findings, in the present study we tested patients with focal ventromedial prefrontal damage and healthy control subjects in personal and impersonal moral dilemmas. Given that patients with ventromedial prefrontal damage may show deficits in decision-making independent of the moral content of the choice options (e.g. Mavaddat *et al.*, 2000; Fellows and Farah, 2003; Fellows, 2006), we also included a set of non-moral dilemmas for comparison purposes. In order to make our results easily comparable with those by Greene and colleagues (Greene *et al.*, 2001), we used the same dilemmas used by these researchers, although translated to Italian.

If medial prefrontal regions are implicated in opposing personal moral violations, then patients with lesions in this region should be more inclined than healthy controls to approve moral violations in personal moral dilemmas. In contrast, no performance difference was expected between patients and controls in impersonal and non-moral dilemmas, in which behavior is deemed to be less dependent on processing in medial prefrontal areas (Greene *et al.*, 2001).

METHODS

Participants

Participants in the present study included 7 brain-damaged patients and 12 healthy individuals. Brain-damaged patients were recruited from the Centro Studi e Ricerche in Neuroscienze Cognitive, Cesena. They were selected on the basis of the location of their lesion evident on CT or MRI scans. Patients were included who had lesion restricted to the ventromedial prefrontal cortex. In all cases lesions were the result of a ruptured aneurysm of the anterior communicating artery. All patients presented with a decline in social interpersonal conduct (e.g. patient 3's wife reported that they were no longer joining their friends to play cards, because he easily got angry and kept screaming to the others), lack of concern for social rules (e.g. during the clinical sessions patient 6 often made comments concerning the physical appearance of female staff members) and emotional blunting (e.g. patients were invariably reported as no longer interested in the personal life of their close relatives). Patients had a mean age of 55 years (s.d. = 6.8), and a mean education of 10 years (s.d. = 5). The control group consisted of 12 healthy individuals who had been matched to patients on the basis of age (mean age = 57.3 years; s.d. = 6.3) and education (mean education = 12.3 years; s.d. = 4; $P > 0.3$ in both cases). The gender variable was balanced across groups. Participants were excluded if they exhibited clinically significant depression, alcohol or drug abuse, epilepsy or other known neurological condition. Participants gave their informed consent to participate in the study according to the Declaration of Helsinki (1991), that was approved by the Ethical Committee of the Department of Psychology, University of Bologna.

Table 1 shows demographic data, time post-injury, lesion side, etiology, lesion description, as well as the results each patient obtained in neuropsychological tests commonly used in clinical practice. All patients showed preserved intellectual skills, as indicated by the scores obtained on the Mini-Mental State Examination (Folstein *et al.*, 1975), the Verbal Judgment Task (i.e. a test see Spinnler and Tognoni, 1987 for normative data) and the Raven Progressive Matrices (Spinnler and Tognoni, 1987). However, patients' neuropsychological profile included moderate problems in

Table 1 Demographic, clinical and lesion data of the two patient groups

VMPFC Patient	Sex	Age at test (years)	Time since lesion (months)	Education	Side of lesion	Etiology	Description of lesion	MMSE score	SRM score*	VJT score*	WMS score	WCST (perseverative responses)*
1	F	62	42	8	L	ACoA Aneurysm	Vm PFC	24	50	50	85	9
2	F	52	30	19	B	ACoA Aneurysm	Vm PFC	27	34	50	80	30
3	M	66	33	5	L	ACoA Aneurysm	Vm PFC	23	50	50	92	18
4	M	56	116	13	R	ACoA Aneurysm	Vm PFC	24	19	35	82	/
5	M	56	30	13	B	ACoA Aneurysm	Vm PFC	27	50	50	85	5
6	M	49	46	8	L	ACoA Aneurysm	Vm PFC	26	5	35	93	5
7	M	46	35	8	R	ACoA Aneurysm	Vm PFC	24	50	50	94	30
mean	/	55	47	10	/	/	/	25	37	46	87	16

Note. M = male, F = female, L = left, R = right, B = bilateral, ACoA = anterior communicating artery, VmPFC = ventromedial prefrontal cortex, MMSE = mini-mental state examination (Cut-off = 24), WMS = Wechsler memory scale (normal score = 100 ± 15), SRM = standard Raven matrices, VJT = verbal judgment task. Scores in percentile value are indicated with a (*). Percentile values <5 are indicative of impaired performance.

memory and executive functions. Specifically, on the Wechsler Memory Scale (Wechsler, 1987), two out of the seven patients (patients 2 and 4) showed scores 1 s.d. below average performance, suggesting mild memory problems. We note, however, that as a group patients exhibited a General Memory index close to normal (they scored 87, where the normal mean and s.d. are 100 and 15, respectively). As for executive function, two of the seven patients (patients 5 and 6) showed impaired performance on the Wisconsin Card Sorting Test (Spinnler and Tognoni, 1987). The remaining patients, however, as well as the group as a whole, attained normal scores in this test.

Lesion location and extent

The lesion analysis was based on computerized axial tomography (CT) data in five of the seven cases, because of the nature of the neurological damage, specifically, ruptured aneurysms that were subsequently clipped. In two of the subjects, magnetic resonance scans were possible. To reconstruct each patient's lesion, the template method developed by Damasio and Damasio (1989) was used. Briefly, the location and extent of each lesion were traced from slices of CT and/or MRI scans onto axial templates on which Brodmann Areas (BAs) are premarked. Although the locations of BAs in these templates are approximate, they are widely accepted in the neuropsychology and neurology communities. Figure 1 shows the extent and overlap of the brain lesions in the brain-damaged patients. The regions with the most extensive damage across the seven patients were the ventromedial prefrontal areas, in particular BA 10, 12, 24 and 32.

Materials

Materials in the present study were 15 personal moral dilemmas, 15 impersonal moral dilemmas and 15 non-moral dilemmas, which had been randomly selected from a battery of 60 dilemmas developed by Greene and colleagues (2001) the complete battery is available at: www.sciencemag.org/cgi/content/full/293/5537/2105/DC1), and translated into Italian.
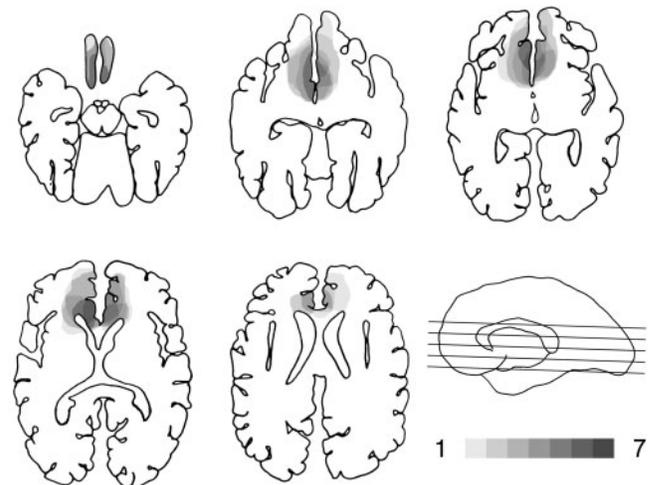
Fig. 1 Location and degree of overlap of brain lesions. The figure shows the lesions of the seven ventromedial prefrontal patients. Lesions are projected on the same five axial templates following the method developed by Damasio and Damasio (1989). The level of the axial slices has been marked by black lines on the mesial view of a right hemisphere drawing. Progressively darker shades denote the degree to which lesions involve the same brain regions, as indicated in the legend.

Moral dilemmas are supposed to elicit moral emotions, while non-moral dilemmas are not (Greene *et al.*, 2001). Typical examples of non-moral dilemmas posed questions about whether to buy a new television or to have your old television repaired for the same price, or whether to travel by bus or train given certain time constraints.

Procedure

Subjects sat in front of the computer. Each dilemma was presented as text through a series of three screens. The first screen described the scenario. The second screen posed a question about the appropriateness of an action one might perform in that scenario, i.e. the 'dilemmatic question' (e.g. 'Is it appropriate to save the five persons by pushing the stranger to death?'). We also added a third screen involving a question about the content of the scenario, i.e. the 'memory question' (e.g. 'Did the number of persons on the main

track equal 10?'). The memory question was introduced in order to make sure that patients were able to remember relevant aspects of the scenario while taking their decisions.

Subjects read at their own pace, pressing a button to advance from one screen to the next. After reading the dilemmatic question subjects responded 'appropriate' or 'inappropriate' by pressing one of two buttons. Participants were told to respond as soon as they had reached a decision. For all dilemmas being tested, 'appropriate' responses implied the maximization of overall consequences (Greene, 2003), e.g. killing one instead of five persons (in a moral dilemma), or buying a new television instead of repairing the old one for the same price (in a non-moral dilemma). However, only for moral dilemmas 'appropriate' responses resulted in moral violations. Note that 'appropriate' and 'inappropriate' is a value-neutral description of what the participant said about the action in the dilemma and not an evaluation of the participant's decision. Both the number of 'appropriate' responses and response times (RTs; i.e. the time from the onset of the dilemmatic question to the moment a response was given) were collected. Once a response was given, the memory question appeared, and participants responded 'YES' or 'NO' by pressing one of two buttons. The proportion of correct responses was taken as a measure of memory accuracy.

Normal subjects received all the 45 dilemmas during a single session, which took about 50 min. In order to reduce fatigue, patients received the 45 dilemmas in three sessions including 15 dilemmas each (five personal, five impersonal and five non-moral dilemmas), and separated by about 3 days. There was no difference across testing sessions in patients' results [$F(2,12) = 0.55$; $P = 0.58$ for the number of 'appropriate' responses; $F(2,12) = 0.84$; $P = 0.45$ for RTs], which were therefore collapsed for the purpose of data analysis.

RESULTS

Preliminary analyses showed that patients were able to remember relevant aspects of the dilemmas' scenario while making their decisions. Indeed, the proportion of correct responses to memory questions was above 0.9 across

subjects, and comparable between patients and controls [0.93 vs 0.95; $t(17) = 1.07$; $P = 0.3$].

Moral vs non-moral
We first compared participants' performance between moral and non-moral judgments. Table 2 shows RTs for 'appropriate' and 'inappropriate' responses. Table 3 shows the proportion of 'appropriate' responses to moral (collapsed across personal and impersonal) dilemmas and non-moral dilemmas.

An ANOVA on RTs with Group (patients, controls) as between-subject factor, and Dilemma (Moral, Nonmoral) and Type of response (appropriate, inappropriate) as within-subject factors revealed a significant Dilemma × Type of response interaction [$F(1,17) = 8.3$; $P < 0.01$]: Participants were slower to give 'appropriate' responses to moral compared to non-moral dilemmas (9191 vs 7142; $P < 0.05$), whereas the opposite trend was observed for 'inappropriate' responses (7262 vs 8539; $P = 0.1$). No other effects were significant ($P > 0.6$ in all cases).

An ANOVA on the proportion of 'appropriate' responses with Group and Dilemma as factors yielded a significant effect of Dilemma [$F(1,17) = 10.2$; $P < 0.005$], indicating that participants gave fewer 'appropriate' responses to moral compared to non-moral dilemmas (0.39 vs 0.57; $P < 0.005$). Group was not significant ($P = 0.4$), and no Group × Dilemma interaction emerged ($P = 0.9$).

This first set of analyses shows that patients (like normal controls) were less inclined and slower to approve moral violations compared to actions with no moral implication. Importantly, patients performed normally in non-moral dilemmas, which reduces the possibility that unspecific deficits in decision-making (e.g. the inability to detect the most advantageous between two choice options) affected our results.

Moral personal vs moral impersonal
We next compared participants' performance between personal and impersonal moral dilemmas. Figure 2 shows RTs for 'appropriate' and 'inappropriate' responses. Figure 3 shows the proportion of 'appropriate' responses to personal and impersonal moral dilemmas.

Table 2 Response time for 'appropriate' and 'inappropriate' responses to moral and non-moral dilemmas in patients and controls

| | Response time (ms) | | | |
| | Moral | | Non-moral | |
	Appropriate	Inappropriate	Appropriate	Inappropriate
Patients	8638	6947	6425	7306
Controls	9744	7577	7860	9773

Table 3 Proportion of 'appropriate' responses to moral dilemmas (collapsed across personal and impersonal) and non-moral dilemmas in patients and controls. Values in parenthesis refer to 1 standard error of the mean

| | Proportion of 'appropriate' responses | |
	Moral	Non-moral
Patients	0.41	0.57
Controls	0.36	0.53
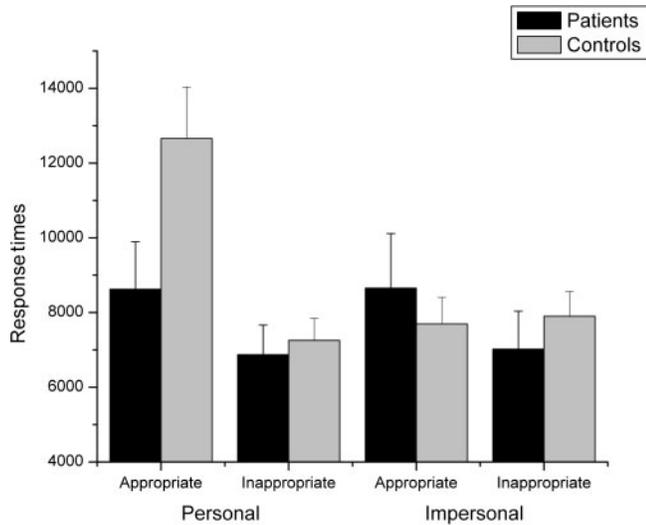
Fig. 2 Response time for 'appropriate' and 'inappropriate' responses to personal and impersonal moral dilemmas in patients and controls. Bars refer to 1 standard error of the mean.
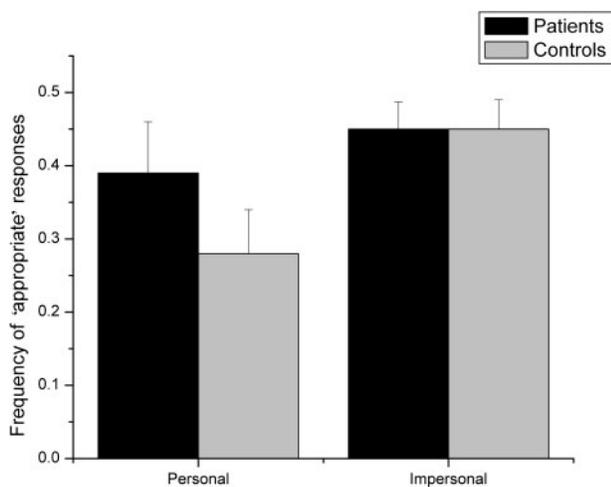
Fig. 3 Proportion of 'appropriate' responses to personal and impersonal moral dilemmas in patients and controls. Bars refer to 1 standard error of the mean.

An ANOVA on RTs, with Group, Dilemma, and Type of response (appropriate, inappropriate) as factors yielded a significant effect of Dilemma $[F(1,15) = 8; P < 0.05]$, which was qualified by a significant Group $\times$ Dilemma interaction $[F(1,15) = 9.3; P < 0.05]$. *Post hoc* comparisons showed that normal controls were slower in giving personal judgments compared with impersonal (10.001 *vs* 7631 ms; $P < 0.01$), whereas in patients both types of judgment took equally long (7748 *vs* 7838 ms; $P = 0.9$). Importantly, a Group $\times$ Dilemma $\times$ Type of response interaction also emerged $[F(1,15) = 5.7; P < 0.05]$: Normal controls took longer to give 'appropriate' compared to 'inappropriate' responses in personal moral dilemmas (12658 *vs* 7345 ms; $P < 0.001$), but not in impersonal moral dilemmas ($P = 0.8$). In contrast, patients showed similar RTs for 'appropriate'

and 'inappropriate' responses in both personal and impersonal dilemmas ($P > 0.5$ in both cases). Interestingly, compared to normal controls, patients were faster to give 'appropriate' responses in personal dilemmas (12658 *vs* 8622; $P < 0.05$) but not in impersonal dilemmas ($P = 0.8$).

An ANOVA on the proportion of 'appropriate' responses with Group and Dilemma (moral personal, moral impersonal) as factors yielded a significant effect of Dilemma $[F(1,17) = 10; P < 0.001]$, such that fewer 'appropriate' responses were given in personal moral dilemmas compared to impersonal (0.33 *vs* 0.44; $P < 0.05$), and no effect of Group ($P = 0.5$). The Group $\times$ Dilemma interaction did not reach statistical significance $[F(1,17) = 2; P = 0.14]$. Nevertheless, for completeness, we also conducted planned comparison. We found that normal controls gave fewer 'appropriate' responses to personal compared to impersonal moral dilemmas (0.28 *vs* 0.45; $P < 0.05$), whereas patients gave a similar proportion of 'appropriate' responses to both types of dilemma (0.39 *vs* 0.45; $P = 0.23$). Compared to normal controls, patients gave more 'appropriate' responses in personal dilemmas (0.28 *vs* 0.39; $P = 0.056$), but the same amount in impersonal dilemmas (0.45).

This set of analyses shows that patients were faster and more inclined then normal controls to authorize moral violations in personal moral dilemmas, whereas their performance in impersonal moral dilemmas was comparable to that of controls. Normal controls were less inclined to approve personal compared to impersonal moral violations, whereas patients showed a comparable behavior when faced with personal and impersonal moral dilemmas.

DISCUSSION

The present study investigated personal and impersonal moral judgment in patients with ventromedial prefrontal lesions and healthy individuals. Given that patients with lesions in the ventromedial prefrontal cortex may show abnormal moral conduct and lack of concern for moral rules (e.g. Saver and Damasio, 1991; Bechara, 2005; Moll *et al.*, 2005), we were interested in verifying whether, and under which conditions, these patients would be more inclined than normal controls to judge moral violations as acceptable behaviors. To this aim, we had patients with ventromedial prefrontal lesions and normal controls consider personal and impersonal moral dilemmas, which required the individual to judge whether it is appropriate or not to incur a moral violation in order to follow utilitarian, more reasoned, considerations (Greene *et al.*, 2001). In moral dilemmas, dissonant moral values of roughly comparable strength strongly conflict, such that no widely accepted formal moral principle exists that establishes *a priori* what behavior is appropriate in these circumstances. Thus, much like in real life, individuals have to decide what they would or would not do based on their on-line appraisal of the specific situation they are contemplating.

Results on healthy individuals replicate and extend those obtained by Greene and colleagues (2001, 2004): Subjects were slower to approve, but relatively quick to condemn, personal moral violations, whereas approvals and disapprovals of impersonal moral violations took equally long. Moreover, subjects approved fewer moral violations in personal moral dilemmas compared to impersonal, and when they did approve a personal moral violation, the decision took longer than for impersonal ones. Patients with ventromedial prefrontal lesions were more inclined to approve personal moral violations compared to normal controls, and did so more quickly. In sharp contrast, their behavior in impersonal and non-moral dilemmas was comparable to that of controls, both in term of the quality of the choices they made, and in the time they needed to make their decisions. Thus, while normal subjects appeared disproportionately reluctant to authorize personal moral violations compared to impersonal ones, patients were as willing to authorize personal as impersonal moral violations.

The evidence of normal behavior in impersonal and non-moral dilemmas suggests that patients' abnormal performance in personal moral dilemmas was not due to unspecific deficits in decision-making consequent to medial prefrontal damage (e.g. Fellows and Farah, 2003), such as an impulsiveness to approve the behaviors called into question without evaluating their merit properly, or even in the cognitive operations supporting decision-making tasks (e.g. maintaining an active representation of the scenario in working memory, shifting attention between the competing behavioral alternatives, etc.; see McKinnon and Moscovitch, 2006 for a discussion). Indeed, such problems would have been apparent in all types of dilemma, for example in a systematic tendency to respond 'appropriate' to the dilemmatic questions, or decrease in RTs across experimental conditions. This was not the case: Patients were able to detect the objectively most advantageous between two behavioral options in non-moral dilemmas and, in doing so, they showed similar RTs to the controls. Also, patients showed increased RTs to moral dilemmas compared to non-moral, suggesting that their behavior was not rigid, but instead reflective of the specific content of the situation being considered.

As well, our results suggest that patients were aware that some actions may or may not be in conflict with moral values and rules while taking their decisions. Indeed, like normal controls, they refused moral violations more frequently compared to actions with no moral implication, as emerged when comparing their behavior in moral and non-moral dilemmas. Note, also, that the time patients needed to refuse moral violations was comparable to that of controls, suggesting that knowledge about moral values was not only available to patients, but also normally accessible. This finding is in line with evidence of retained moral knowledge after lesions in the ventromedial prefrontal cortex (e.g. Saver and Damasio, 1991; Blair and Cipolotti, 2000;

Mendez et al., 2005; but see Anderson et al., 1999, see also Hauser, 2006).

Thus, the results of the present study point to a selective deficit in personal moral judgment in patients with ventromedial prefrontal lesions, in the face of relatively preserved moral knowledge and ability to reason (impersonally) about right and wrong of a complex situation. Similar results have been observed recently in patients with frontotemporal dementia (FTD; Mendez et al., 2005), a progressive neurogenerative disorder that in its early phases may affect ventromedial prefrontal regions disproportionately more than dorsolateral regions (Hodges and Miller, 2001). Patients with FTD were more likely to declare that they would push the stranger to death in the *footbridge dilemma* than were patients with Alzheimer's disease and normal controls, whereas no difference across participant groups was observed in the *trolley dilemma* (Mendez et al., 2005). This finding, again, argues for a specific role of the ventromedial prefrontal cortex in personal moral judgment (Mendez et al., 2005). A potential confound of studies on degenerative disorders is that multiple brain areas may undergo deterioration along with the one of interest, making it difficult to map relations between brain and behavior. In fact, McKinnon and colleagues have recently found contradictory results to those of Mendez in FTD patients, namely a systematic tendency to refuse moral violations in personal moral dilemmas, often based on well-known social dictums (e.g. 'I wouldn't do it because it is wrong to kill'; McKinnon et al., 2006). However, our results on patients with focal lesion in the ventromedial prefrontal cortex clearly reinforce the proposal, advanced by Greene et al. (2001), that this brain region is crucial to oppose personal moral violations. In line with this, several studies have linked criminal behavior to medial prefrontal dysfunction (e.g. Blair, 2001; Kiehl et al., 2001).

It is then natural to ask how the ventromedial prefrontal cortex might accomplish such a role.

As we discussed earlier, moral dilemmas require taking decisions about what is right or wrong to do in a novel situation. Over the course of deciding, individuals tend to consider both the immediate and the long-term consequences of their choices (Bechara, 2005), and to foresee how they would feel about these outcomes (Mellers and McGraw, 2001). Immediate and future prospects may trigger competing signals, whose summation will ultimately shape subjects' choices. It was recently proposed (Bechara, 2005) that during decision-making signals about the immediate and the long-term outcomes of choice options would be conveyed through two separate but interacting neural systems: The amygdala would trigger affective/emotional signals of immediate outcomes, whereas the ventromedial prefrontal cortex would be necessary to trigger affective/emotional signals of long-term ones (Bechara et al., 1994; McClure et al., 2004).

Certainly, in both the *footbridge* and the *trolley dilemma*, blocking the train has appealing immediate consequences: It involves saving five lives, and may convey the sensation of being a hero. On the other hand, individuals must have thought about the negative prospect of causing the death of a man while pondering their decision. Importantly, in normal controls this latter factor, which conflicts with the approval of the moral violation, weighted disproportionately more in personal dilemmas compared to impersonal ones, which was not the case for patients. As Lieberman (2006) recently noted, personal dilemmas may induce individuals to focus on their own personal involvement in bringing about a distasteful outcome, whereas impersonal dilemmas, by definition, lack this sense of agency and responsibility (see also Cushman *et al.*, 2006, Borg *et al.*, 2006). Thus, while contemplating personal dilemmas normal subjects might have anticipated negative emotional responses at the thought of causing direct harm to an individual, such as regret, guilt or an automatic emotional identification with the victim (Greene and Haidt, 2002), which then contributed to decision-making through the ventromedial prefrontal cortex, thereby mediating an aversive reaction to personal moral violations.

We argue that patients' increased tendency to approve personal moral violations related to a failure to anticipate the emotional, self-focused, long-term consequences of their choices (Frijda, 2005; Amodio and Frith, 2006; Tangney *et al.*, 2007). In line with this idea, evidence from several fMRI studies shows that reflecting on one's emotional experience is supported by medial prefrontal regions (e.g. BA10; Gusnard *et al.*, 2001; Ochsner *et al.*, 2004; Eisenberger *et al.*, 2005) which are damaged in our patients. Accordingly, patients with ventromedial prefrontal lesions report reduced self-conscious emotions after engaging in socially inappropriate behaviors compared to patients with dorsolateral lesions (see also Eslinger and Damasio, 1985; Beer *et al.*, 2006). Moreover, these patients may fail to anticipate future emotions (e.g. regret), in order to guide decision-making (Camille *et al.*, 2004). We also note that social reasoning abilities such as empathy heavily rely on processing in medial prefrontal regions (e.g. Brothers and Ring, 1992; Eslinger, 1998), and may be impaired in patients with ventromedial prefrontal damage (Shamay-Tsoory *et al.*, 2005), possibly resulting in reduced responsiveness to victims (see Blair and Cipolotti, 2000).

Other authors have proposed that moral violations in patients with ventromedial prefrontal lesions may result from a failure to develop or use affective/emotional cues to inhibit morally unacceptable behavior (e.g., Damasio, 1994; Hornak *et al.*, 1996; Blair and Cipolotti, 2000; Mah *et al.*, 2005). One might even argue that whether moral judgment results impaired in patients with ventromedial prefrontal lesions would critically depend on the degree to which the task taps emotional/self-focused processing. Thus, whereas ventromedial prefrontal patients are generally able to recognize moral trasgressions as such (this study; Saver and Damasio, 1991; Blair and Cipolotti, 2000; Mendez *et al.*, 2005), they may fail to identify socially inappropriate actions for which there are not formal societal prohibitions but which typically induce negative emotions in observers (e.g. intimately touching another's child), and arguably are detected based on an anticipation of such emotional responses (Blair and Cipolotti, 2000, see also Lough *et al.*, 2006). Similarly, in the study by Beer and colleagues (2006) which we discussed earlier, patients who had previously failed to feel that their behavior was socially inappropriate were able to recognize it as such on a later video recording. This finding, again, suggests that the ventromedial prefrontal cortex would be crucial for self-focused, rather then externally-focused (or even knowledge-driven), social cognition mechanisms (Beer *et al.*, 2006; see also Lieberman, 2006). The dissociation we found in patients between impaired personal moral judgment and preserved impersonal moral judgment provides further support to this interpretation.

In summary, we report that patients with ventromedial prefrontal lesions were more inclined to judge personal moral violations as acceptable than were normal controls. This finding indicates that the ventromedial prefrontal cortex is crucial to oppose personal moral violations. This brain area might be necessary to forecast the long-term emotional consequences of these actions during decision-making (Bechara, 2005), thus preventing their implementation, even when the resulting omission will cause remarkable immediate costs (e.g. in the footbridge dilemma). It has been argued that the ventromedial prefrontal cortex is at the core of a neural mechanism that allows individuals to endure sacrifices now in order to obtain benefits later (Bechara, 2005). We speculate that the long-term benefit of letting five people die in the footbridge dilemma would be to contribute to preserve an overarching moral value (i.e. 'do not kill a member of your own group'), and thus, in turn, the long-term welfare of the community. Similar mechanisms might be at work when individuals punish violations of social fairness at a personal cost (Fehr and Gachter, 2002; Fehr and Fischbacher, 2003). Indeed, fMRI investigations of cooperation and fairness have detected consistent activation of the ventromedial prefrontal cortex (Rilling *et al.*, 2002; Decety *et al.*, 2004). Possibly, we are committed to pass to our successors these pro-social dispositions, that we inherited from our ancestors as a crucial way to strenghten our social group, by means of neural mechanisms that automatically bias the way we take our decisions, independent of our contingent reasons.

REFERENCES

Amodio, D.M., Frith, C.D. (2006). Meeting of minds: the medial prefrontal cortex and social cognition. *Nature Neuroscience Reviews*l, *7*, 268–77.

Anderson, S.W., Bechara, A., Damasio, H., Tranel, D., Damasio, A.R. (1999). Impairment in social and moral behaviour related to early damage in human prefrontal cortex. *Nature Neuroscience*, *2*, 1032–7.

Bechara, A. (2005). Decision making, impulse control and loss of willpower to resist drugs: a neurocognitive prospective. *Nature Neuroscience*, *8*, 1458–63.

Bechara, A., Damasio, A.R., Damasio, H., Anderson, S.W. (1994). Insensitivity to future consequences following damage to human prefrontal cortex. *Cognition*, *50*, 7–15.

Beer, J.S., John, O.P., Scabini, D., Knight, R.T. (2006). Orbitofrontal cortex and social behavior: integrating self-monitoring and emotion-cognition interactions. *Journal of Cognitive Neuroscience*, *18*, 871–9.

Beer, J., Ochsner, K.N. (2006). Social cognition: a multilevel analysis. *Brain Research*, *1079*, 98–105.

Berthoz, S., Armony, J.L., Blair, R.J.R., Dolan, R.J. (2002). An fMRI study of intentional and unintentional (embarassing) violations of social norms. *Brain*, *125*, 1696–708.

Blair, R.J.R. (2001). Neurocognitive models of aggression, the antisocial personality disorders, and psychopathy. *Journal of Neurology, Neurosurgery and Psychiatry*, *71*, 727–31.

Blair, R.J.R., Cipolotti, L. (2000). Impaired social response reversal: a case of 'acquired sociopathy'. *Brain*, *123*, 1122–41.

Borg, J.S., Hynes, C., Van Horn, J., Grafton, S., Sinnot-Armstrong, W. (2006). Consequences, action, and intention as factors in moral judgment: an fMRI investigation. *Journal of Cognitive Neuroscience*, *18*, 803–17.

Brothers, L., Ring, B. (1992). A neuroethological framework for the representation of minds. *Journal of Cognitive Neuroscience*, *4*, 107–18.

Camille, N., Coricelli, G., Sallet, J., Pradat-Diehl, P., Duhamel, J.R., Sirigu, A. (2004). The involvement of the orbitofrontal cortex in the experience of regret. *Science*, *304*, 1167–70.

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*, 1082–9.

Damasio, H., Damasio, A.R. (1989). *Lesion Analysis in Neuropsychology*. New York: Oxford.

Damasio, A.R. (1994). *Descartes' error*. New York: Putnam.

Decety, J., Jackson, P.L., Sommerville, J.A., Chaminade, T., Meltzoff, A.N. (2004). The neural basis of cooperation and competition: an fMRI investigation. *Neuron*, *23*, 744–51.

Declaration of Helsinki (1991). *British Medical Journal*, *302*, 1194.

Eisenberger, N.I., Lieberman, M.D., Satpute, A.B. (2005). Personality from a controlled processing perspective: an fMRI study of neuroticism, extraversion, and self-consciousness. *Cognitive and Affective Behavioural Neuroscience*, *5*, 169–81.

Eslinger, P.J. (1998). Neurological and neuropsychological bases of empathy. *European Journal of Neurology*, *39*, 193–9.

Eslinger, P.J., Damasio, A.R. (1985). Severe disturbance of higher cognition after bilateral frontal lobe ablation: patient EVR. *Neurology*, *35*, 1731–41.

Fellows, L.K. (2006). Deciding how to decide: ventromedial frontal lobe damage affects information acquisition in multi-attribute decision making. *Brain*, *129*, 944–52.

Fellows, L.K., Farah, M.J. (2003). Ventromedial frontal cortex mediates affective shifting in humans: evidence from a reversal learning paradigm. *Brain*, *126*, 1830–36.

Fehr, E., Gachter, S. (2002). Altruistic punishment in humans. *Nature*, *415*, 137–40.

Fehr, E., Fischbacher, U. (2003). The nature of human altruism. *Nature*, *425*, 785–91.

Folstein, M.F., Folstein, S.E., McHugh, P.R. (1975). 'Mini Mental State': a practical method for grading the cognitive state of patients for the clinician. *Journal of Psychiatric Research*, *12*, 189–98.

Frijda, N. (2005). Emotion and experience. *Cognition and Emotion*, *19*, 473–97.

Greene, J. (2003). From neural 'is' to moral 'ought': what are the moral implications of neuroscientific moral psychology? *Nature Reviews Neuroscience*, *4*, 847–50.

Greene, J., Haidt, J. (2002). How (and where) does moral judgment work? *Trends in Cognitive Sciences*, *6*, 517–23.

Greene, J., 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*, 2105–8.

Gusnard, D.A., Akbudak, E., Shulman, G.L., Raichle, M.E. (2001). Medial prefrontal cortex and self-referential mental activity: relation of a default mode of brain function. *Proceedings of the National Academy of Sciences USA*, *98*, 4259–64.

Hauser, M.D. (2006). *Moral Minds*. New York: Harper Collins Publishers.

Heekeren, H.R., Wartenburger, I., Schmidt, H., Schwintowski, H.P., Villringer, A. (2003). An fMRI study of simple ethical decision-making. *Neuroreport*, *14*, 1215–9.

Hodges, J.R., Miller, B. (2001). The neuropsychology of frontal variant frontotemporal dementia and semantic dementia. Introduction to the special topic papers: Part II. *Neurocase*, *7*, 113–21.

Hornak, J., Rolls, E.T., Wade, D. (1996). Face and voice expression identification in patients with emotional and behavioural changes following ventral frontal lobe damage. *Neuropsychologia*, *34*, 247–61.

Kiehl, K.A., Smith, A.M., Hare, R.D., et al. (2001). Limbic abnormalities in affective processing by criminal psychopaths as revealed by functional magnetic resonance imaging. *Biological Psychiatry*, *50*, 677–84.

Lieberman, M.D. (2006). Social cognitive neuroscience: a review of core processes. *Annual Review of Psychology*, *58*, 18.1–18.31.

Lough, S., Kipps, C.M., Treise, C., Watson, P., Blair, R.J.R., Hodges, J.R. (2006). Social reasoning, emotion and empathy in frontotemporal dementia. *Neuropsychologia*, *44*, 950–8.

Mah, L.W., Arnold, M.C., Grafman, J. (2005). Deficits in social knowledge following damage to ventromedial prefrontal cortex. *Journal of Neuropsychiatry and Clinical Neuroscience*, *17*, 66–74.

Mavaddat, N., Kirkpatrick, P.J., Rogers, R.D., Sahakian, B.J. (2000). Deficits in decision-making in patients with aneurysms of the anterior communicating artery. *Brain*, *123*, 2109–17.

McClure, S.M., Laibson, D.I., Loewenstein, G., Cohen, J.D. (2004). Separate neural systems value immediate and delayed monetary rewards. *Science*, *306*, 503–7.

McKinnon, M.C., Levine, B., Moscovitch, M. (2006). Domain-general Contributions to Social Reasoning: the Perspective from Cognitive Neuroscience. In: Roberts, M., editor. *Integrating the Mind: Domain General versus Domain Specific Processes in Higher Cognition*. New York: Psychology Press (in press).

McKinnon, M.C., Moscovitch, M. (2006). Domain-general contributions to social reasoning: theory of mind and deontic reasoning re-explored. *Cognition*, *102*, 179–218.

McKinnon, M.C., Talmi, D., Jaswal, G., et al. (2006). Impairment of moral reasoning performance in frontotemporal dementia. In revision.

Mellers, B.A., McGraw, A.P (2001). Anticipated emotions as guides to choice. *Current Directions in Psychological Science*, *10*, 210–14.

Mendez, M.F., Anderson, E., Shapira, J.S. (2005). An investigation of moral judgement in Frontotemporal Dementia. *Cognitive and Behavioural Neurology*, *18*, 193–7.

Moll, J., de Oliveira-Souza, R., Bramati, I.E., Grafman, J. (2002a). Functional networks in emotional moral and nonmoral social judgments. *Neuroimage*, *16*, 696–703.

Moll, J., de Oliveira-Souza, R., Eslinger, P.J., et al. (2002b). The neural correlates of moral sensitivity: a functional magnetic resonance imaging investigation of basic and moral emotions. *Journal of Neuroscience*, 22, 2730–6.

Moll, J., Zahn, R., de Oliveira-Souza, R., Krueger, F., Grafman, J. (2005). The neural basis of human moral cognition. *Nature Reviews Neuroscience*, 6, 799–809.

Ochsner, K.N., Ray, R., Cooper, J.C., et al. (2004). For better or for worse: neural systems supporting the cognitive down- and up-regulation of negative emotion. *Neuroimage*, 23, 483–99.

Rilling, J., Gutman, D., Zeh, T., Pagnoni, G., Berns, G., Kilts, C. (2002). A neural basis for social cooperation. *Neuron*, 35, 395–405.

Saver, J.L., Damasio, A.R. (1991). Preserved access and processing of social knowledge in a patient with acquired sociopathy due to ventromedial prefrontal damage. *Neuropsychologia*, 29, 1241–9.

Shamay-Tsoory, S.G., Tomer, R., Berger, B.D., Goldsher, D., Aharon-Peretz, J. (2005). Impaired 'affective theory of mind' is associated with right ventromedial prefrontal damage. *Cognitive and Behavioral Neurology*, 18, 55–67.

Spinnler, H., Tognoni, G. (1987). Standardizzazione e Taratura Italiana di Test Neuropsicologici. *The Italian Journal of Neurological Science*, 6(Suppl. 8).

Stuss, D., Gow, C.A., Hetherington, C.R. (1992). 'No longer Cage': frontal lobe dysfunction and emotional changes. *Journal of Consulting and Clinical Psychology*, 60, 349–59.

Tangney, J.P., Stuewig, J., Mashek, D.J. (2007). Moral emotions and moral behavior. *The Annual Review of Psychology*, 58, 1–23.

Wechsler, D, editor (1987). *The Wechsler Memory Scale*. London: Psychological Corporation.