Update TRENDS in Cognitive Sciences Vol.11 No.8

Research Focus

Moral judgments, emotions and the utilitarian brain

Jorge Moll and Ricardo de Oliveira-Souza

Cognitive and Behavioral Neuroscience Unit, LABS-D'Or Hospital Network, Rio de Janeiro, RJ 22281-080, Brazil

The investigation of the neural and cognitive mechanisms underlying the moral mind is of paramount importance for understanding complex human behaviors, from altruism to antisocial acts. A new study on patients with prefrontal damage provides key insights on the neurobiology of moral judgment and raises new questions on the mechanisms by which reason and emotion contribute to moral cognition.

Introduction

What makes people recoil upon witnessing human tragedies and engage in costly helping behaviors or violently protest against acts of injustice? Since the early reports of social behavioral changes following brain injury [1,2], neuroscience has begun to provide crucial evidence bridging brain and morality, mercilessly shaking commonsense beliefs [3]. Functional neuroimaging and brain lesion analysis have espoused sophisticated cognitive models and tools and are fueling rapid advances in our understanding of human morality, which relies on partially overlapping abilities, such as the capacity to make moral judgments and experience moral emotions, and to behave according to moral standards. In this vein, a new study by Koenigs, Young and colleagues [4] provides important evidence that bilateral damage to the ventromedial prefrontal cortex (VMPFC) increases 'utilitarian' choices in moral dilemmas (i.e. judgments favoring the aggregate welfare over the welfare of fewer individuals), strongly supporting the notion that normal moral judgment springs from a complex interaction of cognitive and emotional mechanisms relying on specific neural structures. At the same time, the study raises further questions on the mechanisms by which the VMPFC influences moral judgments.

In the study [4], the performance of six patients with bilateral VMPFC damage (Figure 1a) on moral decision-making tasks was compared with that of patients with other brain lesions and with neurologically normal controls. Moral and non-moral scenarios pertained to four main classes: (i) 'high conflict' emotionally salient 'personal' moral scenarios (e.g. pushing a bulky stranger onto the track of a runaway trolley to save the lives of five workmen, thus killing the stranger); (ii) 'low-conflict' emotionally salient 'personal' scenarios (e.g. hiring a man to rape your wife while you're away so that you can comfort her and conquer her love again); (iii) less emotionally salient, 'impersonal' scenarios (e.g. lying to a guard to borrow a speedboat and warn tourists of an impending storm); and (iv) non-moral scenarios (e.g. take the train instead of the bus to arrive in time). 'Personal' scenarios were framed in such a way that a 'Yes' choice meant accepting a highly aversive moral violation [e.g. saying that it is appropriate to hire the rapist (low-conflict dilemma), or that it is appropriate to push the bulky stranger (high-conflict dilemma)]. VMPFC patients and controls unanimously responded 'no' to the low-conflict 'personal' scenarios. However, VMPFC patients endorsed 'utilitarian' decisions in high-conflict scenarios – highly emotionally aversive choices that would nonetheless lead to greater aggregate welfare (e.g. more lives saved) – much more often than control subjects did. This study provides a direct link between damage to a circumscribed brain region and a change in preferences to emotionally salient moral judgments, a dissociation within the moral judgment domain.

The VMPFC and morality

Several studies have documented changes in social behavior following damage to different cortical and subcortical structures. Such behavioral impairments can vary from social inadequacy (e.g. lack of social tact) to severe moral violations (e.g. pedophilia). Although there is extensive evidence for a role of several brain regions in the implementation and regulation of moral behavior [5], the VMPFC has been focused on the most in this field. This is understandable, given its involvement in several neural mechanisms that, although not specific for morality, are important for the organization of moral behavior. These include outcome prediction, associative learning and flexible evaluation of behavioral contingencies [6]. In a landmark study, impairments in interpersonal behavior were demonstrated in a patient with acquired prefrontal damage who was unimpaired on standard measures of moral reasoning [7]. Early damage to the VMPFC, which often extends to the frontopolar cortex [FPC, Brodmann's area (BA) 10], can lead to severe impairments of both moral behavior and reasoning, suggesting that these prefrontal regions are crucially important in moral learning [8]. More recently, functional imaging studies using a wide range of tasks (such as active moral judgments and passive exposure to morally salient stimuli) began to provide specific evidence for the role of the VMPFC and FPC in moral reasoning and moral emotions in normal adults (Figures 1b–e) [9,10]. Such VMPFC–FPC activations occurred together with activations in the anterior temporal cortex, superior temporal sulcus region and limbic structures, leading to the concept of the 'moral brain rsquo; as a network of closely interconnected regions that integrates the diverse functions involved in moral appraisals [5].

Corresponding author: Moll, J. (mollj@neuroscience-rio.org).
Available online 29 June 2007.

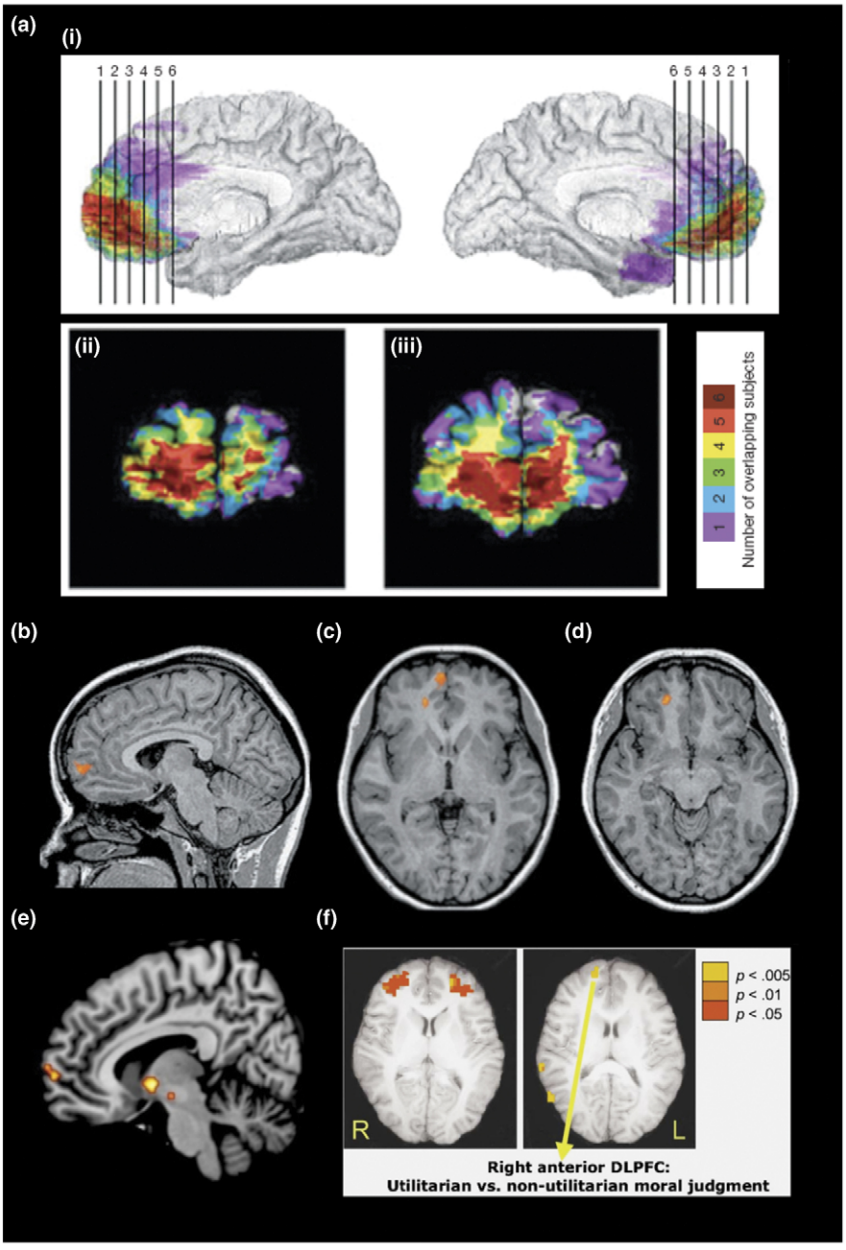

Figure 1. Ventromedial prefrontal cortex (VMPFC) and frontopolar cortex (FPC) in moral judgment and emotion. **(a)** Group lesion mapping in the study by Koenigs *et al.* [4]. (i) Medial views of the right and left hemispheres. (ii,iii) The coronal slices correspond to the two most anterior cuts across the prefrontal region. The VMPFC was damaged in all patients; the overlap between different patients is shown by the coloring. Lesions extended to the medial sector of the FPC in at least five out of six patients, and to more lateral sectors of the FPC (encroaching the anterior DLPFC) in four out of six of them. **(b,c)** FPC activation in response to passive exposure to morally salient pictures as compared to emotionally salient non-moral pictures [9] in (b) sagittal and (c) transversal views. **(d)** Activation of the more ventral medial orbitofrontal cortex in the same study (effects controlled for valence and arousal) [9]. **(e)** FPC and basal forebrain activation during passive presentations of non-dilemmatic scenarios evocative of prosocial moral emotions (conjunction of guilt and compassion, in comparison with emotionally neutral social scenarios) [14]. **(f)** the moral judgment fMRI study in normal subjects by Greene *et al.* [13] of moral judgment fMRI in normal subjects, the most reliable effects in response to utilitarian choices (whole-brain analysis) were observed in the medial FPC (arrow, which was described as anterior DLPFC by the authors), with weaker effects spreading to more lateral sectors (using a region-of-interest analysis) [13]. The medial and lateral areas both correspond to BA 10 and thus are part of the FPC [11]. These regions were damaged in most VMPFC patients of the Koenigs *et al.* study [4] (see (a)). BA, Brodmann's area; DLPFC, dorsolateral prefrontal cortex; fMRI, functional magnetic resonance imaging; FPC, frontopolar cortex; VMPFC, ventromedial prefrontal cortex. Reproduced, with permission, from Refs (a) [4]; (b–d) [9]; (e) [14]; (f) [13].

Reason and emotion in moral judgment: are we there yet?

A central issue in studies on the relationship between morality and brain damage relates to the precise anatomical distribution of the lesions, a point that warrants careful consideration in Koenigs *et al.* study [4]. In the VMPFC group, prefrontal damage extended bilaterally to the medial FPC in five of the six patients (Figure 1a), and to the lateral FPC (encroaching the anterior dorsolateral PFC (DLPFC), BA 9/46) in ~four of them. Other gray and white matter structures belonging to orbital, dorsomedial and cingulate sectors were more variably affected. Because the FPC differs from the VMPFC in cytoarchitectonic, connectional, functional, and behavioral respects [11], damage to

this area could have important implications, which are discussed below in the context of the experimental findings of Koenigs *et al.* [4].

The increased preference of VMPFC–FPC patients for utilitarian choices can be interpreted according to different functional–anatomic hypotheses. Making more 'rational', utilitarian choices in difficult dilemmas might have resulted from a general emotional blunting due to VMPFC damage. However, as discussed by Koenigs *et al.* [4], results from another study by Koenigs *et al.* [12] do not support this hypothesis. In the two-person Ultimatum game, participants had to choose between accepting an unfair but financially rewarding proposal (an economically 'rational' choice), or rejecting it to punish the unfair player (an 'emotional' choice). VMPFC patients opted more often than controls for rejecting unfair offers (i.e. they were more 'emotional').

According to another view, proposed by Greene *et al.* [13], emotion and cognition (or reason) have mutually competing roles in moral judgment [13]. Utilitarian choices in difficult moral dilemmas arise from cognitive control mechanisms based in the DLPFC, whereas non-utilitarian choices emerge from emotional responses relying on the medial PFC. Testing Greene *et al.*'s dual-process view of mutually competing cognitive control and emotion would require, however, the demonstration of a double dissociation – showing both that selective VMPFC damage increases utilitarian choices and also that selective DLPFC or lateral FPC damage leads to emotional choices. Furthermore, the FPC region most robustly activated by utilitarian choices in Greene *et al.*'s functional magnetic resonance imaging (fMRI) study [13], and the most lateral extension of the FPC (located in BA 10; Figure 1f), were damaged in most patients in Koenigs *et al.*'s study (Figure 1a). Therefore, the dual-process hypothesis neither explains Koenigs *et al.*'s findings better than a simple dissociation hypothesis (i.e. an overall impaired emotional experience), nor is supported by the finding of increased emotional choices in the Ultimatum game by VMPFC patients [12].

A third and more parsimonious explanation would be that the VMPFC–FPC might be necessary for the experience of prosocial moral sentiments. It has been proposed that these complex feelings emerge from integration, instead of conflict, between emotional and cognitive mechanisms. The experience of compassion and empathic concern, for example, requires the engagement of limbic-mediated emotional states (e.g. sadness or attachment) in conjunction with mechanisms mediated by the FPC, such as prospective thinking and representing multiple outcomes of events and actions (e.g. forecasting the consequences of our own acts onto others) [5]. Indeed, fMRI studies have shown that the VMPFC and FPC are consistently engaged not only by tasks requiring explicit moral judgments [10], but also by passive presentation of stimuli evocative of moral emotions, in the absence of cognitive conflict or typical executive processes [9,14] (Figure 1b–d). Our research suggests that prosocial moral emotions might rely strongly on medial fronto-limbic networks (Figure 1e), with lateral sectors of the DLPFC and orbitofrontal cortex being more important for self-centered and other-aversive emotional experience (e.g. anger, frustration or moral disgust) [5,9,15]. This hypothesis explains why VMPFC patients are 'more rational' when judging 'personal' moral dilemmas [4] but more emotional in the Ultimatum game [12] and is in line with the decreased empathic concern and guilt of the patients reported by Koenigs *et al.* [4].

The study by Koenigs, Young and colleagues [4] provides novel insights but raises new conundrums for cognitive neuroscience, fuelling the enthusiasm of researchers of the human moral mind. Exploring the cognitive and neural organization of higher-order morally salient representations (i.e. moral sentiments and values) and how they interact with complex cognitive abilities, such as outcome prediction, to guide moral judgments is a fascinating area for future research.

Acknowledgements

We thank Mike Koenigs, Liane Young and Marc Hauser for their helpful comments and clarifications. We also thank Roland Zahn for helpful discussions, and the reviewers for their constructive criticisms. J.M. was supported by LABS-D'Or Hospital Network and by PRONEX/CNPq.

References

1 Welt, L. (1888) Ueber charackterveranderungen des menschen. *Dtsch. Arch. Klin. Med.* 42, 339–390
2 Macmillan, M. (2000) *An Odd Kind of Fame: Stories of Phineas Gage,* MIT Press
3 Damasio, A. (2003) *Looking for Spinoza: Joy, Sorrow and the Feeling Brain,* Harcourt
4 Koenigs, M. *et al.* (2007) Damage to the prefrontal cortex increases utilitarian moral judgements. *Nature* 446, 908–911
5 Moll, J. *et al.* (2005) Opinion: the neural basis of human moral cognition. *Nat. Rev. Neurosci.* 6, 799–809
6 Rolls, E.T. (1996) The orbitofrontal cortex. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 351, 1433–1443
7 Eslinger, P.J. and Damasio, A.R. (1985) Severe disturbance of higher cognition after bilateral frontal lobe ablation: patient EVR. *Neurology* 35, 1731–1741
8 Eslinger, P.J. *et al.* (1992) Developmental consequences of childhood frontal lobe damage. *Arch. Neurol.* 49, 764–769
9 Moll, J. *et al.* (2002) The neural correlates of moral sensitivity: a functional magnetic resonance imaging investigation of basic and moral emotions. *J. Neurosci.* 22, 2730–2736
10 Moll, J. *et al.* (2001) Frontopolar and anterior temporal cortex activation in a moral judgment task: preliminary functional MRI results in normal subjects. *Arq Neuropsiquiatr.* 59, 657–664
11 Christoff, K. and Gabrieli, J.D.E. (2000) The frontopolar cortex and human cognition: evidence for a rostrocaudal hierarchical organization within the human prefrontal cortex. *Psychobiology* 28, 168–186
12 Koenigs, M. and Tranel, D. (2007) Irrational economic decision-making after ventromedial prefrontal damage: evidence from the Ultimatum Game. *J. Neurosci.* 27, 951–956
13 Greene, J.D. *et al.* (2004) The neural bases of cognitive conflict and control in moral judgment. *Neuron* 44, 389–400
14 Moll, J. *et al.* The self as a moral agent: linking the neural bases of social agency and moral sensitivity. *Social Neuroscience* (in press)
15 Moll, J. *et al.* (2006) Human fronto-mesolimbic networks guide decisions about charitable donation. *Proc. Natl. Acad. Sci. U. S. A.* 103, 15623–15628

1364-6613/$ – see front matter
doi:10.1016/j.tics.2007.06.001