Investigating emotion in moral cognition: a review of evidence from functional neuroimaging and neuropsychology

Liane Young†, and Michael Koenigs*‡

†Department of Psychology, Harvard University, 33 Kirkland Street, 984 William James Hall, Cambridge, MA 02138, USA, and‡Cognitive Neuroscience Section, National Institute of Neurological Disorders and Stroke, National Institutes of Health, 10 Center Drive MSC 1440, Bethesda, MD 20892, USA

*These authors contributed equally to this work.

Introduction: Human moral decision-making has long been a topic of philosophical debate, and, more recently, a topic for empirical investigation. Central to this investigation is the extent to which emotional processes underlie our decisions about moral right and wrong. Neuroscience offers a unique perspective on this question by addressing whether brain regions associated with emotional processing are involved in moral cognition.

Method: We conduct a narrative review of neuroscientific studies focused on the role of emotion in morality. Specifically, we describe evidence implicating the ventromedial prefrontal cortex (VMPC), a brain region known to be important for emotional processing.

Results: Functional imaging studies demonstrate VMPC activation during tasks probing moral cognition. Studies of clinical populations, including patients with VMPC damage, reveal an association between impairments in emotional processing and impairments in moral judgement and behaviour.

Conclusions: Considered together, these studies indicate that not only are emotions engaged during moral cognition, but that emotions, particularly those mediated by VMPC, are in fact critical for human morality.

Keywords: Neuroscience/emotion/morality/ventromedial prefrontal cortex

Knowing right from wrong may be as fundamental to human experience as language, vision or memory. Indeed, as Thomas Jefferson once proclaimed, the moral sense of man ‘is as much a part of his nature as the
sense of hearing, seeing, feeling’. Science has taught us a great deal about how these latter cognitive capacities work, in particular, their physical basis in the brain. Human moral cognition, however, has remained largely in the domain of traditional philosophy. For centuries, philosophers, including the likes of Immanuel Kant, David Hume and John Stuart Mill, have debated over how we ought to make moral judgements. For example, should morality be based upon reasons or passions? Should we take consequences into account? Or should we consider instead rights and duties?

Classic moral dilemmas tend to pit such considerations against each other.¹⁻⁵ In one such dilemma, a physician has five patients, each facing his imminent demise due to organ failure of some kind. This physician has another patient who is healthy. The only way that the physician can save the lives of the first five patients is to transplant five of the healthy patient’s organs (against his will) into the bodies of the other five patients. If she does this, the one patient will die, but the other five patients will live. What should she do? In another dilemma, a public health official must decide whether his agency should encourage the use of a recently developed vaccine. The vast majority of people who take the vaccine will develop immunity to a deadly disease, but a small minority will contract the very disease that the vaccine is designed to prevent. What should he do?

Cases like these are standard fare in moral philosophy and can be distilled down to the following question: when, if ever, is it morally permissible to harm one to save many? According to some moral philosophers, our intuitions about individual cases can shed some light on how this question ought to be answered and thus how we ought to behave.⁶ But the empirical study of morality has a different aim, and instead seeks to determine how we do behave, or how we do decide what is right and wrong. Indeed, systematic investigation of how individuals make moral judgements may provide insight into the biologically necessary and sufficient conditions of, or constraints on, the moral mind. For example, what are the biological mechanisms that underlie our answers to moral questions? Why do we favour one response over another? Or, in the case of a particularly sticky moral dilemma, why do we have such difficulty deciding at all? One window into the basis of human moral judgement is the brain.

Here, we review functional magnetic resonance imaging (fMRI) and neuropsychological studies on the neural basis of moral judgement. While important implications of this research could be discussed at length (for example, for law and education⁷⁻⁸), and the term ‘moral judgement’ is itself a topic for rich philosophical debate,⁹⁻¹¹ we will limit the scope of our review to neuroscientific studies investigating the role of emotional processes in moral cognition.
In the past decade, numerous fMRI studies have revealed an association between the processing of morally relevant stimuli and activity in areas of the brain thought to be involved in empathy and emotional responsiveness, in particular, the ventral and medial portions of prefrontal cortex, collectively referred to as ‘ventromedial prefrontal cortex’ (VMPC). Damage to VMPC results in striking impairments in emotional function, including generally blunted affect, diminished empathy, emotional lability and poorly regulated anger and frustration. Furthermore, VMPC projects to limbic, hypothalamic and brainstem regions that execute visceral and autonomic components of emotional responses, and neurons within the VMPC encode the emotional value of sensory stimuli. Thus fMRI activations within this area, which includes the medial portions of orbitofrontal cortex (Brodmann areas 11 and 12) as well as medial prefrontal cortex from the ventral surface to around the level of the genu of the corpus callosum (Brodmann area 25 and portions of Brodmann areas 10 and 32), suggest the engagement of emotional processing.

The use of fMRI in the study of moral cognition began with paradigms contrasting neural responses to moral versus non-moral stimuli. In one such study, subjects viewed emotionally evocative pictures with moral content (e.g. physical assaults, war scenes) and emotionally evocative pictures with non-moral content (e.g. body lesions, dangerous animals). Regions of VMPC, in this case the right medial orbitofrontal cortex and medial frontal gyrus (Brodmann areas 10 and 11; Fig. 1A), were selectively activated during the passive viewing of moral compared with non-moral pictures. Similar VMPC activations (lower medial Brodmann area 10; Fig. 1E) were observed when moral and non-moral pictures were additionally matched for social content and specifically when subjects were asked to regulate their own emotional response.

**Fig. 1** fMRI activations associated with moral cognition. Activations in VMPC are associated with (A) viewing pictures with moral content, (B) viewing statements with moral content, (C) judgments of simple statements with moral content, (D) judgments of moral dilemmas featuring physical harm, and (E) regulation of moral emotions.
response to the moral pictures. The patterns of neural activation in these studies have been interpreted as empirical support for an association between emotional processing and moral cognition, insofar as the passive viewing of morally salient pictures counts as moral cognition.

Another set of studies represents an additional step towards characterizing the relationship between emotion and moral cognition, in using not pictures but ‘moral statements,’ or, more precisely, simple descriptions of morally relevant actions. Activation within VMPC, in this case the left medial orbitofrontal cortex (Fig. 1B), was observed for emotionally salient statements with moral content (e.g. ‘He shot the victim to death’), when compared with emotionally salient non-moral social statements (e.g. ‘He licked the dirty toilet’). A different study stripped the statements in both conditions of explicit emotional content, using moral statements that did not feature direct bodily harm or violence (e.g. ‘A steals a car’/’A admires a car’). Here subjects were asked to judge whether the action was morally appropriate or not. In the control condition, subjects were presented with non-moral statements that were either semantically appropriate (e.g. ‘A takes a walk’) or not (e.g. ‘A waits a walk’). VMPC activation (medial Brodmann area 10; Fig. 1C) emerged during moral as opposed to semantic judgement. A similar study required that subjects make silent ‘right’ or ‘wrong’ judgements of simple moral statements (e.g. ‘We break the law when necessary’) and non-moral statements (e.g. ‘Stones are made of water’). Again, activation within VMPC was specific to the processing of the moral statements.

Convergent evidence from this body of neuroimaging studies suggests at the very least a consistent association between brain areas involved in emotional processing and aspects of moral cognition. However, these studies leave significant questions unanswered. Are emotions associated with the processing not only of images and simple statements with moral content but also of complex moral scenarios? Are emotions associated in a systematic way with specific moral responses under specific conditions?

Greene and colleagues were the first to investigate whether emotion-related areas of the brain are activated during responses to complex moral scenarios, and, moreover, whether these areas are differentially activated for different kinds of moral scenarios, such as ‘personal’ versus ‘impersonal’ scenarios. This distinction contrasts ‘intuitively “up close and personal” (and putatively more emotional)” scenarios with ‘intuitively impersonal (and putatively less emotional)” scenarios. As predicted, selective VMPC activation, in medial Brodmann area 10 (Fig. 1D), was found for ‘personal’ over ‘impersonal’ moral scenarios. Building on this finding, Greene and colleagues tested the further hypothesis that emotion is associated not only with particular moral
scenarios but with particular moral judgements, specifically, ‘non-utilitarian’ judgements—judgements that are based on factors other than the sum consequences of the action in question. Non-utilitarian judgements made in the context of Greene and colleagues’ study took the form of rejecting as ‘inappropriate’ a harmful action that maximized aggregate welfare. Utilitarian judgements consisted of endorsing the harmful action as ‘appropriate’. A sample dilemma from the study follows:

Enemy soldiers have taken over your village. They have orders to kill all remaining civilians. You and some of your townspeople have sought refuge in the cellar of a large house. Outside, you hear the voices of soldiers who have come to search the house for valuables. Your baby begins to cry loudly. You cover his mouth to block the sound. If you remove your hand from his mouth, his crying will summon the attention of the soldiers who will kill you, your child, and the others hiding out in the cellar. To save yourself and the others, you must smother your child to death. Is it appropriate for you to smother your child in order to save yourself and the other townspeople?

The non-utilitarian judgement in this case is the negative response (e.g. ‘Don’t smother the baby’). This judgement might be based on an emotional aversion to the specific act of smothering one’s baby or a more general \textit{a priori} rule dictating that one must do no harm. Greene and colleagues proposed that subjects’ prepotent response is an aversive emotional response to the harmful act, leading to the rejection of the act. Generating the utilitarian response (e.g. ‘The baby will die no matter what; therefore, smother the baby, save everyone else’) requires overcoming the prepotent emotional response. The utilitarian choice therefore represents the emotionally incongruent or ‘cognitive’ response. Support for this proposal emerged in the finding that increased activity in brain areas associated with cognitive conflict and abstract reasoning, such as anterior cingulate and dorsolateral prefrontal cortex, was found in trials in which subjects generated the utilitarian response. In other words, subjects were able to override an emotional aversion to the harmful act in question by engaging in utilitarian reasoning, a process marked by conflict between ‘cognitive’ and emotional processes. Thus, according to Greene and colleagues, moral judgement is best explained by a dual-system model in which distinct ‘cognitive’ and emotional processes compete.

The observed patterns of activation in these neuroimaging studies have been taken as evidence for emotional engagement during moral judgement. But neuroimaging studies succeed in showing only an association between emotional processing and performance on a
diverse set of moral tasks. They do not, given the limits of current technology, settle the question of whether the observed pattern of activation is a cause or consequence of the moral judgements generated. And, accordingly, they do not settle the question of whether emotional processing plays a systematic role in moral judgement, driving some kinds of moral judgements over others.

**Emotional deficits and moral cognition: neuropsychological studies**

One approach to the question of whether emotional processing is merely associated with or in fact plays a causal role in moral cognition is to test individuals with selective deficits in emotional processing. By examining clinical populations, we can determine the extent to which impaired emotional processing leads to impaired moral cognition and therefore the extent to which normal emotional processing is necessary for normal moral cognition. Indeed, patient studies have suggested that deficits in moral behaviour and cognition are typically associated with emotional dysfunction.

One influential study systematically investigated the moral cognition of two adult individuals who had sustained lesions involving VMPC early in their childhoods.22 In a standardized assessment of moral reasoning, both individuals presented justifications of their moral judgements that did not reflect normal adult levels of moral reasoning, according to a traditional characterization of moral development.23 In particular, their justifications suggested an early or ‘preconventional’ stage of moral reasoning, in which moral dilemmas are approached from a largely egocentric perspective of punishment-avoidance. This finding was particularly remarkable because previous neuropsychological research had shown that adult-onset damage to the same brain region was associated with normal adult levels of moral reasoning.24 The moral reasoning defect in the childhood-onset cases suggests that areas of prefrontal cortex are necessary for the original acquisition of the precursors of normal mature justifications.22 It is worth noting, however, that these studies relied on an analysis not of the judgements themselves but of the justifications of the judgements. As a result, the only conclusions to be drawn from this study concern the effect of emotional deficits on consciously reasoned moral justifications rather than moral judgements per se.

Tests of adult and developmental psychopaths have also associated emotional impairment with defects in moral cognition. This clinical population characteristically exhibits pronounced emotional impairment...
(considerably reduced empathy and guilt) and behavioural disturbance (criminal activity, and, frequently, violence).\textsuperscript{25} In a pair of systematic investigations, both adult and developmental psychopaths were found to have difficulty distinguishing between unambiguous moral transgressions (e.g. hitting someone) and unambiguous conventional transgressions (e.g. talking out of turn) along the dimensions of permissibility, seriousness and authority contingency.\textsuperscript{26,27} One limitation, however, is that these tests focus on the ability to distinguish between moral and conventional transgressions rather than the ability to distinguish between right and wrong—indeed, to generate moral judgements. Thus, whether psychopaths would also show abnormal judgements of right and wrong in the context of moral dilemmas remains an open question.

The investigation of moral judgement (as opposed to moral behaviour or moral reasoning) in brain-damaged populations was first approached through the study of patients with frontotemporal dementia (FTD), which involves deterioration of prefrontal and anterior temporal brain areas (Fig. 2). FTD patients exhibit blunted emotion and diminished regard for others early in the disease course. Behavioural changes include moral transgressions such as stealing, physical assault and unsolicited or inappropriate sexual advances.\textsuperscript{28} Mendez and colleagues explored moral judgement in FTD by asking patients to evaluate a pair of hypothetical moral scenarios.\textsuperscript{29} In both scenarios, a trolley is headed for five people, and the subject has a chance to save the five by sacrificing the life of one person. In one scenario (the ‘impersonal’ scenario), the subject can endorse pulling a lever to turn the trolley away from the

\begin{center}
\includegraphics[width=\textwidth]{brain_degeneration.png}
\end{center}

\textbf{Fig. 2} Brain degeneration in FTD. VMPC dysfunction in FTD patients is associated with (A) apathy, indexed by hypometabolism with PET,\textsuperscript{32} (B) disinhibition, indexed by hypoperfusion with SPECT\textsuperscript{33} and (C) impairment in social judgment, indexed by atrophy with VBM.\textsuperscript{34}
five people onto a side track where one person will be hit and killed instead. In the other scenario (the ‘personal’ scenario), the subject can endorse pushing a large stranger off a footbridge onto the tracks below, where his body will stop the trolley from hitting the five, though he of course will be killed. Whereas the typical response from healthy populations to this pair of scenarios is to advocate pulling the lever but not pushing the stranger, most FTD patients approved of both actions. One interpretation is that the emotional salience of the personal harm (e.g. pushing the stranger) did not impact moral judgments of the FTD patients to the same extent as in healthy populations. The authors of the study suggest that this result is driven by the deterioration of emotional processing mediated by VMPC. Since neurodegeneration in FTD affects multiple prefrontal and temporal areas; however, firm structure-function relationships cannot be concluded.

The relationship among emotion, moral judgement and VMPC has been most clearly elucidated through the study of individuals with focal VMPC lesions. Like FTD patients, VMPC lesion patients exhibit blunted affect and diminished empathy, but unlike FTD patients, VMPC lesion patients retain broader intellectual function. Thus, VMPC patients can be studied in order to assess the role of emotion in moral judgement. Koenigs et al. tested a group of six patients with focal, adult-onset, bilateral lesions of VMPC (Fig. 3A) to determine whether emotional processing subserved by VMPC is in fact necessary for normal moral judgement. In this study patients evaluated a series of moral scenarios, some of which featured relatively low-emotion

---

**Fig. 3** Lesion overlaps in moral judgment studies. Two studies of patients with focal VMPC lesions demonstrate impaired moral judgment for emotional moral dilemmas following VMPC damage. (A) VMPC lesion overlap by Koenigs et al. (B) VMPC lesion overlap by Ciaramelli et al.
‘impersonal’ harms, such as lying on a resume to improve career prospects, whereas others featured highly emotionally aversive ‘personal’ harms, such as smothering a crying baby to escape detection and execution by enemy soldiers, as in the scenario described above. As in previous fMRI studies,1,21 in a subset of the personal scenarios an emotionally aversive harm was pitted against the ‘greater good’. VMPC patients responded normally to the impersonal moral scenarios, but for the personal scenarios the VMPC patients were significantly more likely to endorse committing an emotionally aversive harm if a greater number of people would benefit—the utilitarian response. In plain terms, the VMPC patients’ moral judgements appeared to be based more on consideration of the ‘ends’, as opposed to the ‘means’. A second lesion study confirmed this basic finding by Koenigs et al.35,36
Together, these studies suggest that social emotions mediated by VMPC are indeed necessary for certain kinds of moral judgement.

The study of clinical populations has illuminated many aspects of normal cognition. In the case of moral cognition, it appears that emotion plays an integral role, an observation consistent with recent behavioural studies as well.37,38 A coarse summation of the clinical findings is that individuals who exhibit abnormal emotional processing also exhibit systematically abnormal moral judgement.

One must be wary, however, of overextending the logic of neuropsychology. Morality may not lend itself to science in the same way that vision or language or memory does. Even though the acquisition or expression of moral knowledge may be a suitable subject of scientific inquiry, science cannot reveal what is morally right or morally wrong. Scientific data may be able to establish that at least some normal moral judgements represent a product of emotional as opposed to purely ‘rational’ processes, but when it comes to bridging the gap from normal to normative, the relative absence of such ‘emotional’ judgements in the case of patients with VMPC damage is neither here nor there. In other words, before we can even ask questions concerning corrective measures or therapeutic practices, we should return to the likes of Kant, Hume and Mill or join the efforts of a new camp of scholars, empirical philosophers, who seek to marry descriptive and normative approaches to human moral psychology.39–41 Harming one to save many might seem like defective moral reasoning in some cases (recall the physician who removes a healthy patient’s organs to save five other dying patients). Doing so elsewhere, however, as in the case of well developed but imperfect vaccines, appears to be, if not the clearly prudent choice, at least not an unequivocally reprehensible one. The brain may thus constrain the moral mind, but how we decide to deal with such constraints may be best determined in philosophical debate.
Acknowledgements

We are grateful to Ralph Adolphs for his support.

Funding

This work was supported by the National Institute of Neurological Disorders and Stroke intramural research program (M.K.) and the National Science Foundation (L.Y.).

References