How (and where) does moral judgment work?

Joshua Greene and Jonathan Haidt

Moral psychology has long focused on reasoning, but recent evidence suggests that moral judgment is more a matter of emotion and affective intuition than deliberate reasoning. Here we discuss recent findings in psychology and cognitive neuroscience, including several studies that specifically investigate moral judgment. These findings indicate the importance of affect, although they allow that reasoning can play a restricted but significant role in moral judgment. They also point towards a preliminary account of the functional neuroanatomy of moral judgment, according to which many brain areas make important contributions to moral judgment although none is devoted specifically to it.

Why do we care so strongly about what other people do, even when their actions won't affect us? And how do we decide that someone else has done something wrong? These questions are at the heart of moral psychology, and psychologists' answers to these questions have changed with intellectual fashion. Historically, psychologists have disagreed about whether moral judgments are primarily products of emotional and non-rational processes (such as Freudian internalization or behaviorist reinforcement) or of reasoning and 'higher' cognition (as in Piaget's and Kohlberg's post-conventional reasoning). Recently, however, findings from several areas of cognitive neuroscience have begun to converge on an answer: emotions and reasoning both matter, but automatic emotional processes tend to dominate.

Trends in moral psychology

During the cognitive revolution of the 1950s and 1960s, behaviorist and Freudian theories gave way to mental models and information processing as the preferred framework in psychology. In the moral domain, Lawrence Kohlberg was a part of this revolution. He built on the earlier work of Jean Piaget [1] to develop a six-stage model of the development of moral reasoning [2]. According to Kohlberg, moral growth is driven not by simple brain maturation but rather by experience in 'role taking', or looking at a problem from multiple perspectives. Role taking improves moral reasoning, and moral reasoning, Kohlberg thought, drives moral judgment.

But as the cognitive revolution matured in the 1980s, many researchers began calling for a complementary 'affective revolution'. Kohlberg's focus on moral reasoning seemed to ignore the importance of moral emotions. At the same time, new findings in evolutionary psychology [3,4] and primatology [5] began to point to the origins of human morality in a set of emotions (linked to expanding cognitive abilities) that make individuals *care* about the welfare of others (e.g. kin altruism, including feelings of sympathy), and about cooperation, cheating, and norm-following (e.g. reciprocal altruism, including feelings of shame, gratitude and vengeance).

Integrating affect and reasoning

In the 1990s the affective revolution was reinforced by a new focus on 'automaticity' – the mind's ability to solve many problems, including high-level social ones, unconsciously and automatically [6]. A recent comprehensive model, the social intuitionist model [7], brings together research on automaticity with findings in neuroscience and theory in evolutionary psychology. This model suggests that moral judgment is much like aesthetic judgment: we see an action or hear a story and we have an instant feeling of approval or disapproval. These feelings are best thought of as affect-laden intuitions, as they appear suddenly and effortlessly in consciousness, with an affective valence (good or bad), but without any feeling of having gone through steps of searching, weighing evidence, or inferring a conclusion. These intuitions – for example, about reciprocity, loyalty, purity, suffering – are shaped by natural selection, as well as by cultural forces. People certainly do engage in moral reasoning, but, as suggested by studies of informal reasoning [8], these processes are typically one-sided efforts in support of pre-ordained conclusions. (As William James said, '*A great many people think they are thinking when they are merely rearranging their prejudices.*') Moral reasoning matters, but it matters primarily in social contexts in which people try to influence each other and reach consensus with friends and allies.

This emphasis on quick, automatic affective reactions is supported by recent findings in social psychology, such as: (1) that people evaluate others and apply morally laden stereotypes automatically [9]; (2) that motivations to maintain relationships and defend against threatening ideas bias judgments and motivate subsequent reasoning [10,11]; and (3) that people can very easily construct post-hoc reasons to justify their actions and judgments [12–14].

Somatic markers and decision-making

In keeping with this affective trend, Antonio Damasio and colleagues have generated widespread interest in the affective neural bases of social judgment through their ongoing study of patients with damage to the ventral and medial portions of the frontal lobes [15,16]. To varying degrees, these patients resemble

Joshua Greene
Dept of Psychology, Green Hall, Princeton University, Princeton, NJ 08544-1010, USA.
e-mail: jdgreene@princeton.edu

Jonathan Haidt
Dept of Psychology, University of Virginia, P.O. Box 400400, Charlottesville, VA 22904-4400, USA.
e-mail: haidt@virginia.edu

 1364-6613/02/$ – see front matter PII: S1364-6613(02)02011-9

Phineas Gage, the 19th century railroad foreman who made neurological history after an accidental explosion sent a tamping iron through his medial prefrontal cortex, robbing him of his sound judgment and remarkably little else [15,16]. Damasio and colleagues argue that contemporary patients with Gage-like damage (such as patient EVR [17]), have emotional deficits and, more specifically, an inability to generate and effectively use 'somatic markers', neural representations of body states that imbue behavioral options with affective significance and thus guide on-line decision-making. These patients' deficits are revealed in their abnormal skin-conductance responses (an index of autonomic arousal) and poor performance on the Iowa Gambling Task, which simulates real-life decision-making [17]. These deficits exist to a surprising extent against a background of preserved 'cognitive function' as indexed by IQ tests and other measures. Moreover, such patients exhibit preserved abstract social knowledge in spite of their disastrous real-life judgment. Their affective deficits render them unable to feel their way through life, which suggests that normal decision-making is more emotional and less reasoned than many have believed [15].

Frontal damage and anti-social behavior

Frontal patients like EVR (see above) are more likely to hurt themselves than other people, but a recent study by Anderson *et al*. of two patients with prefrontal damage acquired during early childhood reports behavior that is more flagrantly immoral [18,19]. These patients lie, steal, have neglected their children, and at times have been physically aggressive – all without apparent remorse. Both patients perform reasonably well on IQ tests and other standard cognitive measures and perform poorly on the Iowa Gambling Task, but unlike adult-onset patients their knowledge of social/moral norms is deficient. Their moral reasoning appears to be, in the terminology of Kohlberg, 'preconventional', conducted from an egocentric perspective in which the purpose is to avoid punishment. Other tests show that they have a limited understanding of the social and emotional implications of decisions and fail to identify primary issues and generate appropriate responses to hypothetical social situations. Thus, it appears that the brain regions compromised in these patients (ventral, medial, and polar aspects of the prefrontal cortex) include structures that are crucial not only for on-line decision-making but also for the acquisition of social knowledge and dispositions towards normal social behavior.

Other studies have documented impaired social behavior resulting from frontal dysfunction [20]. In a study of 279 Vietnam War veterans, Grafman and colleagues found that patients with ventromedial frontal damage tend towards violence and aggression [21]. Raine and colleagues have found that individuals diagnosed with anti-social personality disorder have reduced prefrontal gray matter and exhibit reduced autonomic responses to the performance of a socially stressful task [22], and a recent popular account documenting a large number of case studies attributes violent criminal behavior to a combination of childhood abuse and frontal damage [23].

Psychopaths exhibit extreme anti-social behavior in the absence of observed brain damage. However, a recent neuroimaging study has shown that psychopaths exhibit less emotion-related neural activity than control subjects while responding to emotionally valenced words [24]. Blair and others have conducted several studies that characterize psychopathy as an affective disorder involving a reduction in empathy [25] and consequent deficits in moral judgment both in and out of the laboratory [26]. Blair observes that psychopaths, unlike frontal patients, are marked by their tendency towards instrumental rather than reactive aggression [27].

Neuroimaging

Responses to moral sentences and pictures

A handful of recent studies have used functional neuroimaging to study moral psychology. In an fMRI study, Jorge Moll and colleagues [28] presented subjects with simple claims, some with moral content ('*They hung an innocent*') and others without moral content ('*Stones are made of water*'). Judgments in response to claims with moral content produced increased activity bilaterally in the frontal pole, in the medial frontal gyrus, right cerebellum, right temporal pole, superior temporal sulcus (STS), left orbitofrontal cortex (OFC), left precuneus, and the posterior globus pallidus. A more recent study by Moll and colleagues [29] compared judgments in response to simple moral claims with judgments in response to unpleasantly valenced non-moral claims with social content, many of which evoke disgust (e.g. '*He licked the dirty toilet*', '*Pregnant women often throw up*'). A direct comparison of these two conditions revealed greater activity in the left medial OFC for the moral condition and greater activation in the left lateral OFC as well as the left amygdala for the non-moral/social condition. These results suggest a functional dissociation between neural networks within the OFC and associated structures that specialize in processing different kinds of social/emotional information relevant (in varying degrees) to moral judgment. A third study by Moll and colleagues [30] found similar neural responses to pictures with moral content (e.g. physical assaults, poor abandoned children). The medial frontal and posterior cingulate regions were also activated in an fMRI study of empathy and forgiveness [31]. (See also Table 1.)

Emotional engagement in 'personal' versus 'impersonal' moral judgments

Whereas Moll and colleagues have investigated moral cognition by distinguishing the effects of moral versus non-moral phenomena, Greene and colleagues [32] have drawn a distinction within the moral domain between 'personal' and 'impersonal' moral judgments (see Box 1). Greene and colleagues scanned subjects

Box 1. Two kinds of moral thinking: personal and impersonal

Suppose a runaway trolley is about to run over and kill five people. Suppose further that you can hit a switch that will divert the trolley onto a different set of tracks where it will kill only one person instead of five. Is it okay to hit the switch? Now, what if the only way to save the five people were to push a large person (larger than yourself) in front of the trolley, killing him but saving the others? Would that be okay?

Most people say 'yes' to the first case and 'no' to the second in spite of the fact that these cases are so similar [a]. Although it is easy to generate (and surprisingly difficult to defend) hypotheses about why one *ought* to treat these cases differently [b], Greene *et al.* have attempted to explain how people *do* in fact arrive at this puzzling pair of conclusions [a]. To explain the difference, they posit a distinction between what they believe are two fundamentally different kinds of moral thinking, drawing on capacities that emerged at different stages of human evolution. On the one hand, moral thinking is driven largely by social-emotional dispositions built on those we inherited from our primate ancestors [c,d]. At the same time, humans have a unique capacity for sophisticated abstract reasoning that can be applied to any subject matter. One might suppose, then, that human moral thinking is not one kind of process, but rather a complex interplay between (at least) two distinct types of processes: domain-specific, social-emotional responses and domain-neutral reasoning processes applied in moral contexts.

With this in mind, Greene and colleagues distinguished between 'personal' and 'impersonal' moral violations and judgments. A moral violation is personal if it is: (i) likely to cause serious bodily harm, (ii) to a particular person, (iii) in such a way that the harm does not result from the deflection of an existing threat onto a different party. A moral violation is impersonal if it fails to meet these criteria. One can think of these criteria for personal harm in terms of 'ME HURT YOU', and as delineating roughly those violations that a chimpanzee can appreciate. The 'HURT' condition picks out roughly the kinds of harm that a chimp can understand (e.g. assault, as opposed to, say, tax evasion). The 'YOU' condition requires that the victim be vividly represented as an individual. Finally, the 'ME' condition captures the notion of 'agency' [b], the idea that the action must spring in a vivid way from an agent's will, must be 'authored' rather than merely 'edited' by an agent.

Pushing someone in front of a trolley meets all three criteria and is therefore personal, whereas diverting a trolley involves merely deflecting an existing threat, removing the crucial sense of agency and therefore making this violation impersonal.

References

a Greene, J.D. *et al.* (2001) An fMRI investigation of emotional engagement in moral judgment. *Science* 293, 2105–2108

b Thomson, J.J. (1986) *Rights, Restitution, and Risk: Essays in Moral Theory*, Harvard University Press

c Flack, J.C. and de Waal, F.B.M. (2000) 'Any animal whatever': Darwinian building blocks of morality in monkeys and apes. In *Evolutionary Origins of Morality* (Katz, L.D., ed.), pp. 1–29, Imprint Academic

d Haidt, J. (2001) The emotional dog and its rational tail: a social intuitionist approach to moral judgment. *Psychol. Rev.* 108, 814–834

using fMRI while they responded to a series of personal and impersonal moral dilemmas as well as non-moral dilemmas, all of which involved complex narratives. They found that responding to personal moral dilemmas, as compared with impersonal and non-moral dilemmas, produced increased activity in areas associated with social/emotional processing: medial frontal gyrus, posterior cingulate gyrus, and bilateral STS (originally labeled 'angular gyrus'). By contrast, impersonal and non-moral dilemmas as compared with personal dilemmas produced increased activity in areas associated with working memory: dorsolateral prefrontal and parietal areas (see Fig. 1). They found comparatively little difference between the impersonal-moral and non-moral conditions, suggesting that impersonal moral judgment has less in common with personal moral judgment than with certain kinds of non-moral practical judgment.

Greene *et al.* carried out an analysis of subjects' reaction times to link these imaging data to behavior. Subjects were slow to approve of personal violations but relatively quick to condemn them. By contrast, approvals and disapprovals took equally long for impersonal moral and non-moral judgments. This pattern is explained by subjects' having to overcome their negative emotional responses when approving of personal moral violations as compared with other, less emotionally charged actions (this might be likened to the pattern of interference observed in the Stroop task).

The neuroanatomy of moral judgment

The functional neuroimaging boom has provided a wealth of information about the neuroanatomy of emotion, social cognition, and other neural processes. These data, combined with the lesion and pathology data above, allow us to interpret the results of the imaging studies described in the previous section and thus broaden and refine our understanding of the 'moral brain'.

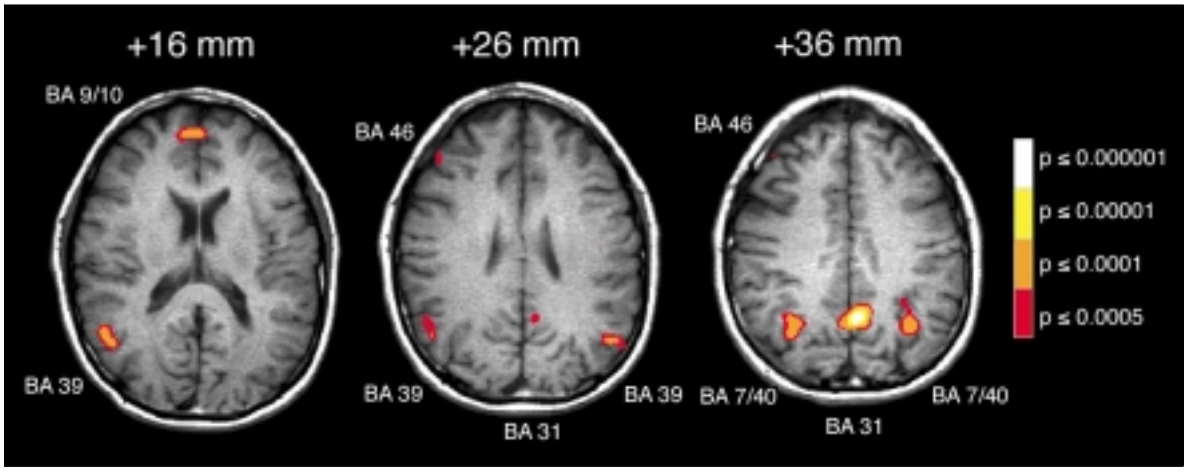

Fig. 1. Brain areas (indicated by Brodmann's area (BA)) exhibiting differences in activity in response to personal moral dilemmas as compared with impersonal and non-moral dilemmas [32]. Areas exhibiting greater activity for personal moral dilemmas (as compared with impersonal and non-moral): medial frontal gyrus (BA 9/10); posterior cingulate gyrus (BA 31); superior temporal sulcus, inferior parietal lobe (BA 39). Areas exhibiting greater activity for impersonal moral dilemmas (as compared with personal): dorsolateral prefrontal cortex (BA 46); parietal lobe (BA 7/40). Images are reversed left to right according to radiologic convention. (Reprinted with permission from Greene *et al.* [32]. Copyright 2001 American Association for the Advancement of Science.)

Table 1. The moral brain (see also Fig. 2, color-coded to this figure.) The first column lists eight brain areas (Brodmann's areas in parentheses) implicated in moral cognition by neuroimaging studies. Subsequent columns provide additional information about their functions.

Brain region (with BA)	Associated moral tasks	Other associated tasks	Social pathology from damage	Likely functions
1. Medial frontal gyrus (BA 9/10)	Personal moral judgments Impersonal moral judgments (relative to non-moral) [32] Simple moral judgments* [28] Viewing moral pictures [30] Forgivability judgments [31]* (*also lateral frontopolar)	Attributing intentionality to moving shapes and cartoon characters* Theory of mind (ToM) stories and cartoons* Representing a historical figure's mental states* [36] Viewing angry/sad faces [47] pleasant pictures, negative pictures [48] (with emotional report [49]) Reward [37] Viewing and/or recall of happy, sad, and disgusting films [50] Emotional autobiographical recall [51] Emotional planning [34] 'Rest' [42] *(focus in the paracingulate sulcus)	Poor practical judgment [15,16], Reactive aggression [27] and (primarily in developmental cases) diminished empathy and social knowledge [18]	Integration of emotion into decision-making and planning [15,16], esp. for conscious processes [33] ToM [36]
2. Posterior cingulate, precuneus, retrosplenial cortex (BA 31/7)	Personal moral judgments Impersonal moral judgments (relative to non-moral) [32] Simple moral judgments [28] Forgivability judgments [31] Moral pictures [30]	Hearing affective autobiographical episodes [52], threat words [38] Reading coherent stories, esp. ToM stories [53] Viewing ToM cartoons [54], familiar faces [55], disgusted faces, sad faces, snake video, previously experienced robbery video, combat pictures (and imagery) Sad autobiographical recall (men) [38] Recognizing neutral words from negative context [56] Emotional planning [34] Recall of happy personal life episodes [57], imaginable word pairs [39] 'Rest' [42]	Impaired recognition memory for faces Capgras delusion? [55]	Integration of emotion, imagery (esp. precuneus [39]), and memory [38], esp. for coherent social narratives
3. Superior temporal sulcus, inferior parietal lobe (BA 39)	Personal moral judgments [32] Simple moral judgments [28,29] Moral pictures [30]	Viewing biological motion (hands, faces, eyes, body) [40]; sad faces [47]; happy, sad, and disgusting films [50,51]; ToM cartoons, reading coherent stories with self-perspective and with characters , esp ToM attributing intentionality to moving shapes Representing a historical figure's mental states [36] Recognizing neutral words from negative context [56] Recall of imaginable word pairs [39] Judgment of indoor/outdoor vs. subjective response to (un)pleasant pictures [49] Emotional film viewing vs. recall [51] 'Rest' [42]	Impaired judgment from eye gaze (monkeys) [40] Capgras delusion? [41]	Supporting representations of socially significant movements [40], and possibly complex representations of 'personhood' [41] ToM [36]
4. Orbitofrontal/ ventromedial frontal cortex (BA 10/11)	Simple moral judgments [28,29] Moral pictures [30]	Reward/punishment [37] Sad autobiographical recall [57] Recognizing words from positive context [56] Viewing angry faces [47] 'Rest' [42] (Note: absent in many PET studies of emotion [34,48–50])	Poor practical judgment [15,16] Reactive aggression [27] and (primarily in developmental cases) diminished empathy and social knowledge [18] Difficulty with advanced ToM tasks [58]	Representation of reward/punishment value [15,16,37] control of inappropriate/ disadvantageous behavior [15,27] 'hot' ToM [58]

The ventral and medial prefrontal cortices

Table 1 and Fig. 2 summarize results relevant to the neuroanatomy of moral judgment. One brain area of great interest is the medial frontal gyrus, around the border of Brodmann areas (BA) 9 and 10, which probably serves in the integration of emotion into decision-making and planning [33,34] and might also play a role in theory of mind (ToM, the capacity to represent others' mental states) and other specifically social functions relevant to moral judgment [35,36]. This

Table 1. continued

Brain region (with BA)	Associated moral tasks	Other associated tasks	Social pathology from damage	Likely functions
5. Temporal pole (BA 38)	Simple moral judgments [28,29]	Reading coherent stories (with characters), ToM stories Attributing intentionality to moving shapes and cartoon characters, Representing a historical figure's mental states Recall of familiar faces and scenes [36] Hearing affective autobiographical episodes [52] Recognition of emotional pictures [59] Viewing emotional pictures (with subjective report) [49], angry/sad faces [47] Viewing and recall of happy, sad (viewing only), and disgusting film [50] Emotional autobiographical recall [51]	Impaired autobiographical memory [59]	Imparting affective tone to experience and memory [59] ToM [36]
6. Amygdala	Moral pictures [30]	Recognition of emotional pictures [59] Viewing emotional film [51], sad faces [47] Viewing racial outgroup faces [44,45]	Poor social judgment from faces and movement [43]	Rapid assessment of reward/punishment value, esp. visual and negative [43]
7. Dorsolateral prefrontal cortex (BA 9/10/46); 8. Parietal lobe (BA 7/40)	Impersonal moral judgment [32]	Working memory and other 'cognitive' tasks [46]		Working memory and other 'cognitive' functions [46]

area should be distinguished from other medial frontal areas, such as the anterior cingulate cortex and a more superior medial frontal region (BA 8/9) that has been implicated in ToM and self-referential tasks [36], but not (yet) in specifically moral tasks. This region should also be distinguished from orbitofrontal/ventromedial areas that appear to be primarily involved in the on-line representation of reward and punishment value [15,17,37]. Although undoubtedly crucial to social behavior and development, the contributions made by these areas may not be specifically social (but see [29]). Rather these regions might perform a more general regulative function in which affective information guides approach and avoidance behavior in both social and non-social contexts. It is worth noting that many of the lesion patients who exhibit acquired sociopathic behavior, such as Gage [16] and EVR [17], have medial damage extending dorsally into BA 9/10. It is also worth noting that many studies that use complex emotional stimuli and PET (which, unlike fMRI, is well-suited to imaging ventral frontal regions) find no activation in orbitofrontal/ventromedial regions [33].

The posterior cingulate and STS

Another region implicated in moral judgment is the posterior cingulate/retrosplenial cortex region centered around BA 31 and the neighboring precuneus area

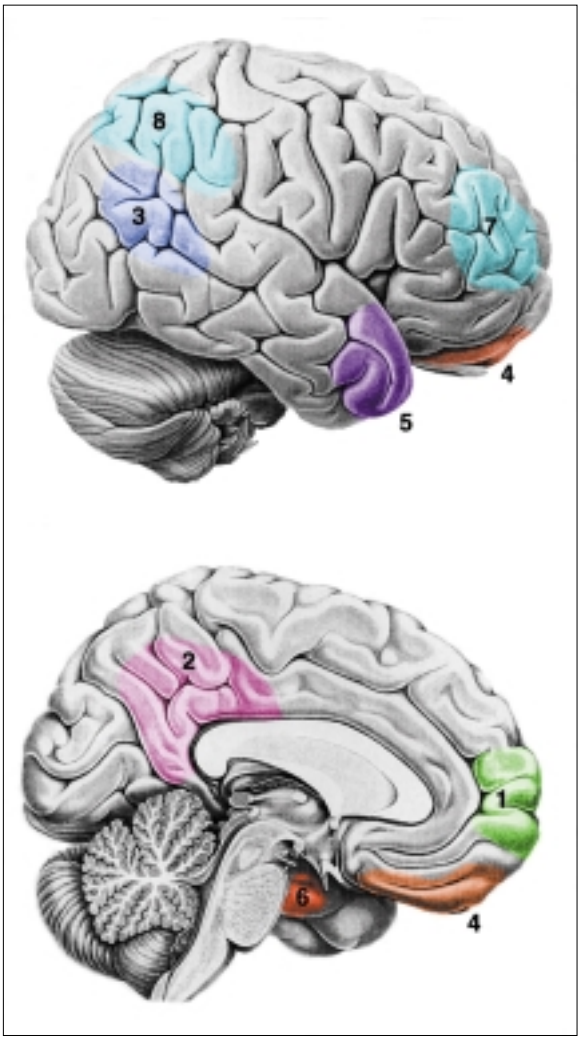

Fig. 2. The moral brain (see also Table 1, color-coded to this figure.) Brain areas implicated in moral cognition by neuroimaging studies (Brodmann's areas in parentheses): 1. medial frontal gyrus (9/10); 2. posterior cingulate, precuneus, retrosplenial cortex (31/7); 3. superior temporal sulcus, inferior parietal lobe (39); 4. orbitofrontal, ventromedial frontal cortex (10/11); 5. temporal pole (38); 6. Amygdala; 7. dorsolateral prefrontal cortex (9/10/46); 8. parietal lobe (7/40). (Adapted with permission from Adolphs (in press) [60].)

Questions for future research

- What are the neural substrates and behavioral effects of specific moral emotions, such as compassion, anger, disgust, guilt and shame? And what about the positive moral emotions that are triggered by the good deeds of others?
- What is the neural basis for cultural and individual variation in moral attitudes? How and when does culture shape moral development?
- How do moral judgments of real events differ from those of the hypothetical stories that have been used for convenience in neuroimaging studies?

extending caudally and dorsally into BA 7. The posterior cingulate/retrosplenial cortex is one of the most commonly activated areas in neuroimaging studies of emotion [38] and one of the areas exhibiting decreased activity in psychopaths during the processing of emotionally valenced words [24]. It appears to be involved in memory and processes involving imagery, especially affective imagery in the context of coherent narrative. The precuneus also appears to be a crucial area for imagery [39].

The posterior STS/inferior parietal region has been an area of intense interest in the emerging field of social cognitive neuroscience for some time. Its primary function appears to be the perception and representation of socially significant information, particularly from biological motion cues, which are crucial for making inferences about the beliefs and intentions of others [36,40]. Brothers and Ring [41] have suggested that more basic social representations of movement (voices, faces, etc.) might serve as building blocks for more complex representations of 'persons'. They note the connection between temporal lobe damage and Capgras delusion, in which patients report that loved ones have been replaced by identical looking and sounding imposters. To the Capgras patient, the faces and voices of loved ones are immediately recognizable, but somehow unfamiliar. What is missing, it seems, is the activation of a high-level, affectively significant representation of the person in question. The suggestion that the posterior STS/inferior parietal region supports representations of 'personhood' dovetails nicely with the finding that this region responds strongly to personal moral dilemmas, and no more strongly to impersonal moral dilemmas than to non-moral dilemmas. (In other words, this region may be responsible for putting the 'YOU' in 'ME HURT YOU' – see Box 1).

The moral brain and the resting brain

In a remarkable convergence, the three brain regions highlighted above coincide with regions identified in a meta-analysis of the resting brain's activity [42]. These regions appear to be components in a tonically active neural system, the activity of which is attenuated when people are engaged in 'goal-directed actions'. Gusnard and Raichle propose that this system's function is the ongoing evaluation of the environment and one's relation to it [42]. We tentatively propose that the key element behind this convergence is introspection; the high-level social-emotional processing involved in moral judgment may be a 'turbo-charged' version of the personal ruminations in which we engage when otherwise unengaged.

The amygdala and dorsolateral 'cognitive' areas

The amygdala plays a crucial role in social-emotional processing [43] and is known to exhibit increased activity in response to racial outgroup faces [44,45] and moral pictures [30]. It appears, however, to be relatively stimulus-driven and biased towards the visual [43], suggesting that its influence on moral judgment is likely to be rather crude. Less crude influences might come from classically 'cognitive' areas in the dorsolateral prefrontal cortex and parietal lobes [46], which have been implicated in impersonal moral judgment [32]. These activations may represent the application of domain-neutral reasoning to moral judgment (see Box 1).

Conclusions

Neuroimaging studies of moral judgment in normal adults, as well as studies of individuals exhibiting aberrant moral behavior, all point to the conclusion, embraced by the social intuitionist model [7], that emotion is a significant driving force in moral judgment. The work of Greene *et al.* [32], however, suggests that reasoning can play an important role in the production of impersonal moral judgments and in personal moral judgments in which reasoned considerations and emotional intuitions conflict. These results also suggest that much, although not necessarily all, moral judgment makes use of processes specifically dedicated to social cognition and, more specifically, the representation of others' mental states (ToM).

We might summarize the likely relationships among social psychological processes, dedicated social cognition, emotion and moral judgment as follows: moral psychology is part of social psychology, but some social psychological processes (e.g. representing another's belief about their physical environment) are not moral. Some social psychological processes appear to make use of cognitive mechanisms specifically dedicated to processing social information (ToM, etc.), and it is likely that some, but not all, moral judgments fall in this category. Some emotions are more central to our moral lives than others (e.g. compassion, guilt and anger), but all emotions can contribute to moral judgment under some circumstances.

Thus, the interrelationships among these overlapping concepts is complicated, and many relevant details remain unclear. What is becoming increasingly clear, however, is that there is no specifically moral part of the brain. Every brain region discussed in this article has also been implicated in non-moral processes. Some experimental designs implicitly suggest a search for the moral part of the brain, however. Moll *et al.* [28], for example, factored subjective ratings of emotional valence into their imaging analysis to 'exclude the effect of emotion on the activation results', as if the emotional processing that occurs during moral judgment must be subtracted out to reveal the truly moral activity.

(The STS and medial frontal activity reported above are from this study's unadjusted results.) In a follow-up study [29] Moll *et al.* compared moral judgments with non-moral emotional judgments to determine 'which brain regions [are] distinctively recruited by emotion processing as opposed to moral judgment.' Such comparisons may be useful as means of distinguishing among the various processes that potentially contribute to moral judgment, but if one attempts to 'deconfound' moral judgment with everything that is not specific to moral judgment (emotion, theory of mind, mental imagery, abstract reasoning, and so on) there will almost certainly be nothing left.

Morality is probably not a 'natural kind' in the brain. Just as the ordinary concept of memory refers to a variety of disparate cognitive processes (working memory, episodic memory, motor memory, etc.), we believe that the ordinary concept of moral judgment refers to a variety of more fine-grained and disparate processes, both 'affective' and 'cognitive'.

Acknowledgements
The authors thank Leigh Nystrom and Andrea Heberlein for their valuable input, and the John Templeton Foundation for its support of J.H.

References

1 Piaget, J. (1965) *The Moral Judgement of the Child*, Free Press
2 Kohlberg, L. (1969) Stage and sequence: the cognitive-developmental approach to socialization. In *Handbook of Socialization Theory and Research* (Goslin, D.A., ed.), pp. 347–480, Rand McNally
3 Trivers, R.L. (1971) The evolution of reciprocal altruism. *Q. Rev. Biol.* 46, 35–57
4 Pinker, S. (1997) *How the Mind Works*, W.W. Norton
5 Flack, J.C. and de Waal, F.B.M. (2000) 'Any animal whatever': Darwinian building blocks of morality in monkeys and apes. In *Evolutionary Origins of Morality* (Katz, L.D., ed.), pp. 1–29, Imprint Academic
6 Bargh, J.A. and Chartrand, T.L. (1999) The unbearable automaticity of being. *Am. Psychol.* 54, 462–479
7 Haidt, J. (2001) The emotional dog and its rational tail: a social intuitionist approach to moral judgment. *Psychol. Rev.* 108, 814–834
8 Galotti, K.M. (1989) Approaches to studying formal and everyday reasoning. *Psychol. Bull.* 105, 331–351
9 Devine, P.G. (1989) Stereotypes and prejudice: their automatic and controlled components. *J. Pers. Soc. Psychol.* 56, 5–18
10 Chen, S. *et al.* (1996) Getting at the truth or getting along: accuracy- versus impression-motivated heuristic and systematic processing. *J. Pers. Soc. Psychol.* 71, 262–275
11 Tetlock, P.E. *et al.* (2000) The psychology of the unthinkable: taboo trade-offs, forbidden base rates, and heretical counterfactuals. *J. Pers. Soc. Psychol.* 78, 853–870
12 Gazzaniga, M.S. (1985) *The Social Brain*, Basic Books
13 Kunda, Z. (1990) The case for motivated reasoning. *Psychol. Bull.* 108, 480–498
14 Nisbett, R.E. and Wilson, T.D. (1977) Telling more than we can know: verbal reports on mental processes. *Psychol. Rev.* 84, 231–259
15 Damasio, A.R. (1994) *Descartes' Error: Emotion, Reason, and the Human Brain*, Putnam
16 Damasio, H. *et al.* (1994) The return of Phineas Gage: clues about the brain from the skull of a famous patient. *Science* 264, 1102–1105
17 Bechara, A. *et al.* (1996) Failure to respond autonomically to anticipated future outcomes following damage to prefrontal cortex. *Cereb. Cortex* 6, 215–225
18 Anderson, S.W. *et al.* (1999) Impairment of social and moral behavior related to early damage in human prefrontal cortex. *Nat. Neurosci.* 2, 1032–1037
19 Dolan, R.J. (1999) On the neurology of morals. *Nat. Neurosci.* 2, 927–929
20 Grattan, L.M. and Eslinger, P.J. (1992) Long-term psychological consequences of childhood frontal lobe lesion in patient DT. *Brain Cogn.* 20, 185–195
21 Grafman, J. *et al.* (1996) Frontal lobe injuries, violence, and aggression: a report of the Vietnam head injury study. *Neurology* 46, 1231–1238
22 Raine, A. *et al.* (2000) Reduced prefrontal gray matter volume and reduced autonomic activity in antisocial personality disorder. *Arch. Gen. Psychiatry* 57, 119–129
23 Pincus, J.H. (2001) *Base Instincts: What Makes Killers Kill?*, W.W. Norton
24 Kiehl, K.A. *et al.* (2001) Limbic abnormalities in affective processing by criminal psychopaths as revealed by functional magnetic resonance imaging. *Biol. Psychiatry* 50, 677–684
25 Blair, R.J. *et al.* (1997) The psychopathic individual: a lack of responsiveness to distress cues? *Psychophysiology* 34, 192–198
26 Blair, R.J. (1995) A cognitive developmental approach to mortality: investigating the psychopath. *Cognition* 57, 1–29
27 Blair, R.J. (2001) Neurocognitive models of aggression, the antisocial personality disorders, and psychopathy. *J. Neurol. Neurosurg. Psychiatry* 71, 727–731
28 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
29 Moll, J. *et al.* (2002) Functional networks in emotional moral and nonmoral social judgments. *Neuroimage* 16, 696–703
30 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
31 Farrow, T.F. *et al.* (2001) Investigating the functional anatomy of empathy and forgiveness. *Neuroreport* 12, 2433–2438
32 Greene, J.D. *et al.* (2001) An fMRI investigation of emotional engagement in moral judgment. *Science* 293, 2105–2108
33 Reiman, E.M. (1997) The application of positron emission tomography to the study of normal and pathologic emotions. *J. Clin. Psychiatry* 58(Suppl 16), 4–12
34 Partiot, A. *et al.* (1995) Brain activation during the generation of non-emotional and emotional plans. *Neuroreport* 6, 1397–1400
35 Castelli, F. *et al.* (2000) Movement and mind: a functional imaging study of perception and interpretation of complex intentional movement patterns. *Neuroimage* 12, 314–325
36 Frith, U. (2001) Mind blindness and the brain in autism. *Neuron* 32, 969–979
37 O'Doherty, J. *et al.* (2001) Abstract reward and punishment representations in the human orbitofrontal cortex. *Nat. Neurosci.* 4, 95–102
38 Maddock, R.J. (1999) The retrosplenial cortex and emotion: new insights from functional neuroimaging of the human brain. *Trends Neurosci.* 22, 310–316
39 Fletcher, P.C. *et al.* (1995) The mind's eye – precuneus activation in memory-related imagery. *Neuroimage* 2, 195–200
40 Allison, T. *et al.* (2000) Social perception from visual cues: role of the STS region. *Trends Cogn. Sci.* 4, 267–278
41 Brothers, L. and Ring, B. (1992) A neuroethological framework for the representation of other minds. *J. Cogn. Neurosci.* 4, 107–118
42 Gusnard, D.A. and Raichle, M.E. (2001) Searching for a baseline: functional imaging and the resting human brain. *Nat. Rev. Neurosci.* 2, 685–694
43 Adolphs, R. (1999) Social cognition and the human brain. *Trends Cogn. Sci.* 3, 469–479
44 Phelps, E.A. *et al.* (2000) Performance on indirect measures of race evaluation predicts amygdala activation. *J. Cogn. Neurosci.* 12, 729–738
45 Hart, A.J. *et al.* (2000) Differential response in the human amygdala to racial outgroup vs ingroup face stimuli. *Neuroreport* 11, 2351–2355
46 Smith, E.E. and Jonides, J. (1997) Working memory: a view from neuroimaging. *Cognit. Psychol.* 33, 5–42
47 Blair, R.J. *et al.* (1999) Dissociable neural responses to facial expressions of sadness and anger. *Brain* 122, 883–893
48 Lane, R.D. *et al.* (1999) Common effects of emotional valence, arousal and attention on neural activation during visual processing of pictures. *Neuropsychologia* 37, 989–997
49 Lane, R.D. *et al.* (1997) Neural activation during selective attention to subjective emotional responses. *Neuroreport* 8, 3969–3972
50 Lane, R.D. *et al.* (1997) Neuroanatomical correlates of happiness, sadness, and disgust. *Am. J. Psychiatry* 154, 926–933
51 Reiman, E.M. *et al.* (1997) Neuroanatomical correlates of externally and internally generated human emotion. *Am. J. Psychiatry* 154, 918–925
52 Fink, G.R. *et al.* (1996) Cerebral representation of one's own past: neural networks involved in autobiographical memory. *J. Neurosci.* 16, 4275–4282
53 Fletcher, P.C. *et al.* (1995) Other minds in the brain: a functional imaging study of 'theory of mind' in story comprehension. *Cognition* 57, 109–128
54 Gallagher, H.L. *et al.* (2000) Reading the mind in cartoons and stories: an fMRI study of 'theory of mind' in verbal and nonverbal tasks. *Neuropsychologia* 38, 11–21
55 Shah, N.J. *et al.* (2001) The neural correlates of person familiarity. A functional magnetic resonance imaging study with clinical implications. *Brain* 124, 804–815
56 Maratos, E.J. *et al.* (2001) Neural activity associated with episodic memory for emotional context. *Neuropsychologia* 39, 910–920
57 Damasio, A.R. *et al.* (2000) Subcortical and cortical brain activity during the feeling of self-generated emotions. *Nat. Neurosci.* 3, 1049–1056
58 Stone, V.E. *et al.* (1998) Frontal lobe contributions to theory of mind. *J. Cogn. Neurosci.* 10, 640–656
59 Dolan, R.J. *et al.* (2000) Dissociable temporal lobe activations during emotional episodic memory retrieval. *Neuroimage* 11, 203–209
60 Adolphs, R. Physiologie und Anatomie der Emotionen. In *Handbuch der Neuropsychologie* (Karnath, H-O. and Thier, P., eds), Chapter 47, Springer-Verlag (in press)