The Cognitive Neuroscience of Moral Judgment*

Joshua D. Greene

To appear in

The Cognitive Neurosciences IV

*This is not the final version, but it is pretty similar

Department of Psychology

Harvard University

33 Kirkland St.

Cambridge, MA 02138

617-495-3898

jgreene@wjh.harvard.edu

This article reviews recent advances in the cognitive neuroscience of moral judgment. The field began with studies of individuals who exhibit abnormal moral behavior, including neurological patients and psychopaths. Such studies continue to provide valuable insights, particularly concerning the role of emotion in moral decision-making. Recent functional neuroimaging studies of normal individuals have identified neural correlates of specific emotional processes relevant to moral judgment. A range of studies using diverse methods support a dual-process theory of moral judgment according to which utilitarian moral judgments (favoring the "greater good" over individual rights) are enabled by controlled cognitive processes, while deontological judgments (favoring individual rights) are driven by intuitive emotional responses. Several recent neuroimaging studies focus on the neural bases of mental state attribution in the context of moral judgment. Finally, research in the field of neuroeconomics has focused on neural processing related to cooperation, trust, and fairness.

The aim of cognitive neuroscience is to understand the mind in physical terms. This endeavor assumes that the mind *can* be understood in physical terms, and, insofar as it is successful, validates that assumption. Against this philosophical backdrop, the cognitive neuroscience of moral judgment takes on special significance. Moral judgment is, for many, the quintessential operation of the mind beyond the body, the Earthly signature of the soul (Greene, in press). (In many religious traditions it is, after all, the quality of a soul's moral judgment that determines where it ends up.) Thus, the prospect of understanding moral judgment in physical terms is especially alluring, or unsettling, depending on your point of view.

In this brief review I provide a progress report on our attempts to understand how the human brain makes moral judgments. In recent years, we have continued to learn valuable lessons from individuals whose abnormal brains dispose them to abnormal social behavior. We have developed new moral-psychological testing materials and used them to dissociate and characterize the affective and cognitive processes that shape moral decisions. Finally, the field of neuroeconomics has brought a welcome dose of ecological validity to the study of moral decision-making. I discuss each of these developments below. (Important and relevant developments in other fields, such as animal behavior and developmental psychology (de Waal, 2006; Hamlin et al., 2007; Warneken et al., 2007), are beyond the scope of this article.)

Bad brains

In the 1990s, Damasio and colleagues published a series of path-breaking studies of decision-making in patients with damage to ventromedial prefrontal cortex (VMPFC), one of the regions damaged in the famous case of Phineas Gage (Damasio, 1994; Macmillan, 2000). VMPFC patients were mysterious because their real-life decision-making was clearly impaired by their lesions, but their deficits typically evaded detection using standard neurological measures of executive function (Saver and Damasio, 1991). Notably, such patients showed no sign of impairment on Kohlberg's (Colby and Kohlberg, 1987) widely used test of moral reasoning (Anderson et al., 1999). Using a game designed to simulate real-world risky decision-making (The Iowa Gambling Task), Bechara and colleagues (1996) documented these behavioral deficits and demonstrated, using autonomic measures, that these deficits are emotional. It seems that such patients make poor decisions because they are unable to generate the feelings that guide adaptive decision-making in healthy individuals.

A later study targeting moral judgment (Anderson et al., 1999) compared patients with adult-onset VMPFC damage to two patients who had acquired VMPFC damage as young children. While the late-onset patients make poor real-life decisions (e.g. neglecting relatives and friends, involvement in shady business ventures), indicating a deterioration of “moral character” (Damasio, 1994), their behavior tended to harm themselves as much as others. The early-onset patients, however, developed into “sociopathic” adults who, in addition to being irresponsible and prone to risk-taking, are duplicitous, aggressive, and strikingly lacking in empathy. What’s more, these two patients, unlike the late-onset patients, exhibited a child-like “preconventional” pattern of moral judgment, reasoning about moral issues from an egocentric perspective focused on reward and punishment. This result suggests a critical role for emotion in moral development. The late-onset patients are prone toward bad decision-making, but, thanks to a lifetime of emotional experience, they are not truly sociopathic. The early-onset patients, in contrast, lacked the emotional responses necessary to learn the basics of human moral behavior. (See also Grattan & Eslinger (1992))

Studies of psychopaths and other individuals with anti-social personality disorder (APD) underscore the importance of emotion in moral decision-making. APD is a catch-all diagnosis for individuals whose behavior is unusually anti-social. Psychopathy, in contrast, is a more specific, somewhat heritable (Blonigen et al., 2005; Viding et al., 2005) disorder whereby individuals exhibit a pathological degree of callousness, lack of empathy or emotional depth, and lack of genuine remorse for their anti-social actions (Hare, 1991). Psychopaths tend to engage in instrumental aggression, while other individuals with APD are characterized by reactive aggression (Berkowitz, 1993; Blair, 2001).

Psychopathy is characterized by profound, but selective emotional deficits. Psychopaths exhibit normal electrodermal responses to threat cues (e.g. a picture of shark's open mouth), but reduced responses to distress cues (e.g. a picture of a crying child) (Blair et al., 1997). In a particularly revealing study, Blair (1995) demonstrated that psychopaths fail to distinguish between rules that authorities cannot legitimately change ("moral" rules, e.g. a classroom rule against hitting) from rules that authorities can legitimately change ("conventional" rules, e.g. a rule prohibiting talking out of turn). According to Blair, psychopaths see all rules as *mere* rules because they lack the emotional responses that lead ordinary people to imbue moral rules with genuine, authority-independent moral legitimacy.

Findings concerning the specific neural bases of psychopathy and APD are varied, implicating a wide range of brain regions including the orbital frontal cortex (OFC)/VMPFC, insula, anterior cingulate cortex (ACC), posterior cingulate cortex (PCC), amygdala, parahippocampal gyrus, and superior temporal gyrus (Kiehl, 2006; Raine and Yang, 2006; Muller et al., 2008). Blair (2004; 2007) has proposed that psychopathy arises primarily from amygdala dysfunction, which is crucial for stimulus-reinforcement learning (Davis and Whalen, 2001), and thus for normal moral socialization (Oxford et al., 2003). The amygdala exhibits reduced activity in psychopaths both in non-moral contexts (e.g., in response to emotional words (Kiehl et al., 2001)) as well in socio-moral contexts (e.g. during cooperative behavior in a prisoner's dilemma game (Rilling et al., 2007)). Consistent with this view, Yang and colleagues (2006) found reduced amygdala volume in psychopaths. The VMPFC, which is known to work in concert with the amygdala (Diergaarde et al., 2005; Schoenbaum and Roesch, 2005), also exhibits many of these effects and appears to play a role in (mis)representing the value of behavioral outcomes in psychopathy (Blair, 2007).

A broader suite of brain regions have been implicated in APD (Raine and Yang, 2006), suggesting, among other things, more general deficits in prefrontal function (Raine et al., 1994). These may be due to structural abnormalities involving reduced prefrontal gray matter (Raine et al., 2000; Yang et al., 2005). There is some evidence implicating abnormal function in dorsolateral prefrontal cortex (DLPFC) in patients with APD (Schneider et al., 2000; Vollm et al., 2004), but not in patients with psychopathy. Given the DLPFC's role in cognitive control (Miller and Cohen, 2001), this is consistent with the notion that psychopaths' aggression results from lack of empathy for others, rather than poor impulse control.

Mapping moral emotion

Consistent with research on APD, and in keeping with a broader trend in moral psychology (Haidt, 2001), most research using functional imaging to study morality has focused on mapping the "where" and "when" of moral emotion in the brain. Some early studies compared moral and non-moral stimuli (Moll et al., 2001; Moll et al., 2002a; Moll et al., 2002b) and identified a suite of brain regions that are sensitive to moral stimuli including the OFC, mPFC, frontal pole, PCC/precuneus, superior temporal sulcus (STS), and temporal pole. This approach, while informative, depends critically on the choice of non-moral control stimuli and the assumption that the observed results are in some way specific to morality (Greene and Haidt, 2002). More recent functional imaging studies have focused on identifying and functionally characterizing different kinds of moral-emotional processes.

Empathy, caring, and harm: Greene and colleagues (2001) identified a set of brain regions associated with judging actions involving "personal," as compared to "impersonal," harm: mPFC (BA 9/10), the PCC/Precuneus (BA 31), and the posterior superior temporal sulcus (pSTS)/temperoparietal junction (TPJ)/angular gyrus (BA 39). (See Figure 1 and "Dual-process morality" below for more details.) A study replicating these results using a larger sample (Greene et al., 2004) identified the same effect in the amygdala, among other regions. The aforementioned regions are implicated in emotional processing (Maddock, 1999; Phan et al., 2002; Adolphs, 2003) as well as in "theory of mind" (ToM) (Frith, 2001; Adolphs, 2003; Amodio and Frith, 2006; Young et al., 2007). These regions, with the exception of the amygdala, are also part of the "default network," (Gusnard and Raichle, 2001; Fox et al., 2005), a set of brain regions that exhibits relatively high levels of tonic activity and that reliably decreases in activity during outwardly directed tasks. Parts of this network are also implicated in self-referential processing (Gusnard et al., 2001; Kelley et al., 2002), episodic memory, "prospection" (Buckner and Carrol, 2007; Schacter et al., 2007), and mind-wandering (Mason et al., 2007). A persistent theme among these processes is the representation of events beyond the observable here and now, such as past, future, and imagined events and mental states. Thus, the activity observed in this network during the contemplation of dilemmas involving "personal" harm is probably related to the fact that these stimuli involve such non-sensory representations, although this alone does not explain why "personal" dilemmas engage this network more than "impersonal" ones. Consistent with this idea, the functional imaging studies of moral judgment that have most robustly engaged this network involve more complex, text-based, narrative stimuli (Greene et al., 2001; Greene et al., 2004; Schaich Borg et al., 2006; Robertson et al., 2007; Young et al., 2007; Greene et al., 2008b; Greene et al., 2008d; Schaich Borg et al., 2008; Young and Saxe, 2008; Kedia et al., in press).

Several studies have focused on neural responses to different types of harm. Luo and colleagues (2006) found that the right amygdala and left VMPFC are sensitive to the intensity of harm displayed in pictures, and Heekeren et al. (2005) found that the amygdala exhibits increased activity in response to narratives involving bodily harm. An earlier study (Heekeren et al., 2003) found no effects in the amygdala using stimuli devoid of violence. Finally, individuals with high psychopathy scores exhibited decreased amygdala activity during the contemplation of moral dilemmas involving "personal" harm (Glenn et al., 2008). Thus, the evidence from functional imaging suggests that the amygdala plays an important role in triggering emotional responses to physically harmful actions.

Specific emotions: Several studies of moral judgment have focused on specific moral emotions, including moral disgust (Rozin et al., 1994; Rozin et al., 1999; Wheatley and Haidt, 2005). Moll and colleagues (Moll et al., 2005) identified a number of brain regions sensitive to core/pathogen disgust in moral contexts. A more recent study (Schaich Borg et al., 2008) compared disgust in response to incestuous acts to pathogen disgust and moral transgressions of a non-sexual nature. Stimuli describing non-sexual moral transgressions (e.g. lying, cheating, stealing), as compared to pathogen-disgust stimuli, elicited increased activity in the familiar mPFC/PCC/TPJ network, but also in the frontal pole/DLPFC and the ACC. Incest descriptions, as compared to that of non-sexual moral transgressions, elicited increased activity in the mPFC/PCC/TPJ network and other regions, including the inferior frontal gyrus, the left insula, the ventral and dorsal ACC, the left amygdala, and the caudate nucleus. Perhaps surprisingly (Phillips et al., 1997; Calder et al., 2001), the insula was preferentially activated only in the incest condition.

Other studies have focused on social emotions such as guilt, embarrassment, shame, pride, empathy, and anger. Robertson and colleagues (2007) found that stimuli associated with care-based morality, as compared to justice-based morality, elicited greater activity in the mPFC and OFC, while the reverse effect was observed in the intraparietal sulcus. In a notably clever study, Beer and colleagues (2003) observed that patients with OFC damage exhibited inappropriate lack of embarrassment when given an opportunity to disclose personal information, inappropriate embarrassment when over-praised for an unremarkable performance on a simple task, and inappropriate pride and lack of embarrassment when describing the nickname they had invented for the experimenter. Berthoz and colleagues (2006) found that the amygdala is especially responsive to the evaluation of intentional transgressions committed (hypothetically) by oneself (guilt), while Kedia and colleagues (in press) observed increased activity in the mPFC, precuneus, and TPJ for evaluations of transgressions involving others (guilt, anger, compassion). These researchers also observed increase activity in the amygdala, ACC, and basal ganglia for transgressions in which both the self and another are involved (guilt, anger). (See also Shin et al. (2000), Berthoz et al. (2002), and Takahashi et al. (2004)).

Moral emotion in context: Other studies have examined the contextual modulation of moral emotion. King and colleagues (2006) used a custom-designed video game to examine violent vs. compassionate behavior in situations in which the behavior is either appropriate (harming an aggressive enemy, helping a distressed innocent person) or inappropriate (helping an aggressive enemy, harming a distressed innocent person). They found that appropriate behavior (whether violent or compassionate) was associated with increased activity in the amygdala and VMPFC. Harenski and Hamann (2006) found that subjects who consciously down-regulated their moral emotions in several lateral regions of PFC. Finally, Finger and colleagues (2006) found that stimuli describing moral/social transgressions committed in the presence of an audience elicited increased activity in the amygdala, underscoring the importance of the amygdala in the social regulation of transgressive behavior (Blair, 2007).

Dual-process morality

The research described above emphasizes the role of emotion in moral judgment (Haidt, 2001), while traditional theories of moral development emphasize the role of controlled cognition (Kohlberg, 1969; Turiel, 2006). I and my collaborators have developed a dual-process theory (Kohlberg, 1969; Posner and Snyder, 1975; Chaiken and Trope, 1999; Lieberman et al., 2002; Kahneman, 2003) of moral judgment that synthesizes these perspectives. (See Figure 2) According to this theory, both intuitive emotional responses and more controlled cognitive responses play crucial and, in some cases, mutually competitive roles. More specifically, this theory associates controlled cognitive processing with utilitarian (or consequentialist) moral judgment aimed at promoting the "greater good" (Mill, 1861/1998). In contrast, this theory associates intuitive emotional processing with deontological judgment aimed at respecting rights, duties, and obligations (Kant, 1785/1959) that may trump the greater good.

We developed this theory in response to a longstanding philosophical puzzle known as the Trolley Problem (Foot, 1978; Thomson, 1985; Fischer and Ravizza, 1992): First, consider the following moral dilemma (which we'll here call the *switch* case (Thomson, 1985)): A runaway trolley is about to run over and kill five people, but you can save them by hitting a switch that will divert the trolley onto a side track, where it will run over and kill only one person. Here, most people say that it is morally acceptable to divert the trolley (Petrinovich et al., 1993), a judgment that accords well with the utilitarian perspective emphasizing the greater good. In the contrasting *footbridge* dilemma, a runaway trolley once again threatens five people. Here, the only way to save the five is to push a large person off a footbridge and into the trolley's path, stopping the trolley but killing the person pushed. (You're too small to stop the trolley yourself.) Here, most people say that it's wrong to trade one life for five, consistent with the deontological perspective, according to which individual rights often trump utilitarian considerations.

We hypothesized that people tend to disapprove of the action in the *footbridge* dilemma because the harmful action in that case, unlike the action in the *switch* case, elicits a prepotent negative emotional response that inclines people toward disapproval (Figure 2e). We hypothesized further that people tend to approve of the action in the *switch* case because, in the absence of a countervailing prepotent emotional response, they default to a utilitarian mode of reasoning that favors trading one life for five (Figure 2a). We proposed that the negative emotional response elicited by the *footbridge* case is related to the more "personal" nature of the harm in that case. We proposed, in other words, there is an emotional appraisal process (Scherer et al., 2001) that distinguishes personal dilemmas like the *footbridge* case from impersonal dilemmas like the *switch* case (Figure 2d).

To test these hypotheses we devised a set of "personal" dilemmas modeled loosely on (and including) the *footbridge* case and a contrasting set of "impersonal" dilemmas modeled loosely on (and including) the *switch* case.[1] The effects of these stimuli were compared using fMRI. As predicted, the personal dilemmas preferentially engaged brain regions associated with emotion, including the mPFC, PCC, and the amygdala (Greene et al., 2001, 2004). (As noted above, this contrast also revealed preferential engagement of the pSTS/TPJ). Also consistent with our dual-process theory, the impersonal moral dilemmas, relative to "personal" ones, elicited increased activity in regions of DLFPC associated with working memory (Cohen et al., 1997; Smith and Jonides, 1997) and cognitive control (Miller and Cohen, 2001).

According to the dual-process theory, the *footbridge dilemma* elicits a conflict between utilitarian reasoning and emotional intuition, where the latter tends to dominate. In other cases these opposing forces appear to be more balanced. Consider the *crying baby* dilemma: It's wartime. You and your fellow villagers are hiding from nearby enemy soldiers in a basement. Your baby starts to cry, and you cover your baby's mouth to block the sound. If you remove your hand, your baby will cry loudly, and the soldiers will hear. They will find you, your baby, and the others, and they will kill all of you. If you do not remove your hand, your baby will smother to death. Is it morally acceptable to smother your baby to death in order to save yourself and the other villagers?

Here, people are relatively slow to respond and exhibit no consensus in their judgments (Greene et al., 2004). According to the dual-process theory, these behavioral effects are the result of the aforementioned conflict between emotional intuition and controlled cognition. This theory makes two key predictions. First, if dilemmas like *crying baby* elicit response conflict, then we would expect these dilemmas (as compared to personal dilemmas that elicit shorter RTs and less disagreement) to be associated with increased activity in the ACC, a region known for its sensitivity to response conflict (Botvinick et al., 2001). (See Figure 2c) Second, if making utilitarian judgments in such cases requires overriding a prepotent, countervailing emotional response, then we would expect such judgments to be associated with increased activity in regions of DLPFC associated with cognitive control (Greene et al., 2001, 2004). (See Figure 2b.) Both of these predictions were confirmed (Greene et al., 2004).

Three more recent studies support the dual-process theory by indicating a causal relationship between emotional responses and deontological/non-utilitarian moral judgments. Mendez et al. (2005) found that patients with frontotemporal dementia, who are known for their "emotional blunting," were disproportionately likely to approve of the action in the *footbridge* dilemma. Koenigs et al. (2007) and Ciaramelli et al. (2007) observed similar results in patients with emotional deficits due to VMPFC lesions. The results of the former study, which distinguished high-conflict personal dilemmas such as the *crying baby* dilemma from low-conflict personal dilemmas, were particularly dramatic. (See Figure 3.) In each of the high-conflict dilemmas, the VMPFC patients gave more utilitarian judgments than the control subjects. Finally, Valdesolo & DeSteno (2006) found that normal participants were more likely to approve of the action in the *footbridge* dilemma following a positive emotion induction aimed at counteracting negative emotional responses.

Four others studies support the link between utilitarian judgment and controlled cognition. My colleagues and I conducted a cognitive load study in which subjects responded to high-conflict personal dilemmas while performing a secondary task (detecting presentations of the number "5" in a stream of numbers) designed to interfere with controlled cognitive processes. The cognitive load manipulation slowed down utilitarian judgments, but had no effect on RT for deontological/non-utilitarian judgments, consistent with the hypothesis that utilitarian judgments, unlike deontological judgments, are preferentially supported by controlled cognitive processes (Greene et al., 2008a). Three other studies have examined the relationship between moral judgment and individual differences in cognitive style. Bartels (2008) found that individuals who are high in "need for cognition" (Cacioppo et al., 1984) and low on "faith in intuition" (Epstein et al., 1996) were more utilitarian. Along similar lines, Hardman (2008) examined moral judgment using the Cognitive Reflection Test (CRT) (Frederick, 2005), which asks people questions like this: "A bat and a ball cost $1.10. The bat costs one dollar more than the ball. How much does the ball cost?" The intuitive answer is 10¢, but a moment's reflection reveals that the correct answer is 5¢. The people who correctly answered these questions were about twice as likely to give utilitarian responses to the *footbridge* and *crying* baby dilemmas. Finally, Moore and colleagues (2008) found that individuals with higher working memory capacity were more likely to give utilitarian judgments in response to dilemmas in which harm to the victim is inevitable. Note however that Kilgore et al. (2007) found that sleep-deprivation made subjects more utilitarian. This effect, however, was not observed in individuals high in emotional intelligence, suggesting the operation of complex emotion-cognition interactions that are not readily explained by current theory.

Three more recent fMRI studies support and broaden the dual-process theory. My colleagues and I compared dilemmas like the *switch* case to similar dilemmas in which saving the five requires breaking a promise. (E.g., the agent had promised the potential victim that he will not be run over.) In these cases it is the harm's *social structure*, rather than its *physical structure*, that generates the tension between utilitarian and deontological judgment. We found, first, that introducing the promise factor reproduces the familiar pattern of mPFC/PCC/TPJ activity and, second, that utilitarian judgment in the promise dilemmas is associated with increased activity in the DLPFC (right BA 46) (Greene et al., 2008b). In a second study (Greene et al., 2008d), we compared utilitarian and non-utilitarian/deontological moral disapproval. The *footbridge* dilemma typically elicits deontological disapproval ("It's wrong to kill the one, despite the greater good,"). One can generate utilitarian disapproval using dilemmas like the *reverse switch* case, in which one can divert the trolley onto five people in order to save one person (an action that makes no utilitarian sense). Consistent with the dual-process theory, we found that utilitarian disapproval, as compared to deontological disapproval, was associated with greater activity in the same region of DLPFC as above. It is worth noting that the region of DLPFC associated with utilitarian judgment in these studies (BA 46) is posterior to that associated with utilitarian judgment in response to high-conflict personal moral dilemmas (BA 10) (Greene et al., 2004). All utilitarian judgments appear to require utilitarian reasoning, but additional cognitive control is only required in the face of countervailing emotional responses. Thus, it is possible that BA 46 is engaged during utilitarian moral reasoning, while BA 10 is engaged in the more extended cognitive processing elicited by high-conflict personal dilemmas. Finally, as noted above, Glenn and colleagues (2008) found that individuals with high psychopathy scores exhibited reduced amygdala activity during the contemplation of personal moral dilemmas, thus providing further evidence for the connection between emotion and deontological impulses (which are reliably generated by personal moral dilemmas). They also found that individuals with high scores on the interpersonal factor of the Psychopathy Checklist (which involves manipulation, conning, superficiality, and deceitfulness) (Hare, 2003) exhibited decreased activation in the mPFC/PCC/TPJ network. (Note, however, that the psychopaths did not exhibit abnormal moral judgment behavior, complicating this interpretation.)

In sum, the dual-process theory of moral judgment, which emphasizes both emotional intuition and controlled cognition, is supported by multiple fMRI studies using different behavioral paradigms, multiple behavioral studies of neurological patients, and a variety of behavioral studies using both experimental manipulations and individual difference measures. (For an alternative perspective see Moll & de Oliveira-Souza (2007). For my reply, see Greene (2007a).)

The mental states of moral agents

As Oliver Wendell Holmes Jr. famously observed, even a dog knows the difference between being tripped over and being kicked. Holmes' comment highlights the importance of information concerning the mental states of moral agents and, more specifically, the distinction between intentional and accidental harm.

Berthoz and colleagues (2006) identified several brain regions that exhibit increased activity in response to intentional (vs. accidental) moral transgressions, including the amygdala, the precuneus, the ACC, and the DLPFC. These results suggest a kind of dual-process response to intentional harms. That may be correct, but a more recent set of studies complicates this picture. Young and colleagues (2007) compared neural responses to intended harms, accidental harms, failed attempted harms, and ordinary harmless actions (a 2 x 2 design crossing mental state information (agent did /did not anticipate harm) and outcome information (harm did/did not result)). They found that that the mPFC, PCC, and TPJ, all regions associated with theory of mind (Saxe et al., 2004), were not only sensitive to belief (i.e. anticipation) information, but were also sensitive to the interaction between belief and outcome information. More specifically, the right TPJ was particularly sensitive to attempted harm, consistent with the behavioral finding that attempted harm is readily condemned, while accidental harm is not so readily excused. (See also Cushman et al. (2006).) Interestingly, Young and colleagues found that judgments in response to accidental harm (as compared to intentional harm) were associated with increased activity in the ACC and DLPFC, regions associated respectively with conflict and control in the context of moral judgment (Greene et al., 2004). Young and colleagues argue that this is due to a conflict between an outcome-based response (the person caused harm) and one based on mental states (it was an accident). Thus, we see here increased activity suggestive of cognitive conflict and control in response to accidental harms, as opposed to intentional harms (Berthoz et al., 2006).

Further studies by Young & Saxe have examined the roles of various neural regions in processing mental state information in the context of moral judgment. They have found that the mPFC is sensitive to the valence of the agent's belief, while the TPJ and precuneus appear to be critical for the encoding and subsequent integration of belief information in moral judgment (Young and Saxe, 2008). A third study (Young and Saxe, in press) suggests that the right TPJ, PCC and mPFC are involved in the generation of spontaneous mental state attributions. Finally, they found, as predicted, that disrupting activity in the right TPJ using TMS produces a more child-like pattern of moral judgment (Piaget, 1965) based more on outcomes and less on mental state information (Young et al., 2008).

While most humans (and perhaps some canines) are explicitly aware of the distinction between intended and accidental harm, people's judgments are also sensitive to a more subtle distinction between harms that are intended as a means to an end and harms that are merely foreseen as side-effects (Aquinas, unknown/2006). The means/side-effect distinction is, in fact, a key distinction that distinguishes the *footbridge* dilemma (in which a person is used as a trolley-stopper) from the *switch* dilemma (in which the a person is killed as "collateral damage") (Foot, 1978; Thomson, 1985; Mikhail, 2000; Cushman et al., 2006; Moore et al., 2008). (Recent research suggests that the means/side-effect distinction interacts with factors related to “personalness” in generating the effects that give rise to the Trolley Problem (Greene et al., 2008c).) Schaich-Borg and colleagues (Schaich Borg et al., 2006) found that the anterior STS and VMPFC exhibit increased activity in response to dilemmas in which the harm is an intended means, as opposed to a foreseen side-effect. They also found increased DLPFC activity associated with harms caused through action, as opposed to inaction, consistent with the finding that people appear to have conscious access to the action/inaction distinction (Cushman et al., 2006).

While the studies described above highlight the importance of mental state representation in moral judgment, a study of moral judgment in autistic children indicates some basic moral judgments do not depend on theory of mind abilities (Leslie et al., 2006).

Neuroeconomics

Morality, broadly construed, may be viewed as a set of psychological adaptations that allow individuals to reap the benefits of cooperation (Darwin, 1871/2004). In economics, the most widely-used experimental paradigm for studying cooperation is the prisoner’s dilemma (Axelrod and Hamilton, 1981), in which two individuals maximize their total payoffs by cooperating, but maximize their individual payoffs by not cooperating (“defecting”). Rilling and colleagues found that brain regions associated with reward (nucleus accumbens, caudate nucleus, VMPFC and OFC, and rostral ACC) were associated with cooperation, indicating that cooperative behavior is supported by general-purpose reward circuitry. A more recent study (Moll et al., 2006) in which people made charitable donations from inside the scanner teaches a similar lesson. Decisions to make costly donations were associated with increased activity in reward-related brain regions overlapping with those identified by Rilling and colleagues. This study also found a remarkably high correlation (*r* = .87) between self-reported engagement in voluntary activities and the level of activation in the mPFC during costly donation.

Several neuroeconomic experiments have used the Ultimatum Game (UG) (Guth et al., 1982) to examine neural responses to unfairness. In the UG, one player (the proposer) makes a proposal about how to divide a fixed sum of money between herself and the other player (the responder). The responder may either accept the proposal, in which case the money is divided as proposed, or reject it, leaving both players with nothing. Responders typically reject offers substantially below half of the total as unfair. Sanfey and colleagues (2003) found that responders responded to such unfair offers with increased activity in the insula, which is associated with autonomic arousal (Critchley et al., 2000) and negative emotion (Calder et al., 2001). The level of insula activity scaled with the magnitude of the unfairness, responded more to human- vs. computer-generated proposals, and was associated with higher levels of rejection. Unfair offers also elicited increased activity in the right DLPFC, which was interpreted as involved in inhibiting the negative emotional response to unfairness. A more recent study (Knoch et al., 2006), however, challenges this interpretation, finding that disrupting activity in the right DLPFC generated fewer rejections of unfair offers. These results suggest that the right DLPFC is involved in inhibiting the appetitive desire for more money, rather than the punitive response to unfair treatment. Koenigs et al. (2007) found that patients with VMPFC damage exhibited the opposite behavioral pattern, suggesting that the VMPFC plays a critical role in regulating the emotional response that drives individuals to respond punitively to unfair treatment. (Increased rejection rates can also be generated by decreasing seratonin levels through tryptophan-depletion (Crockett et al., 2008).) A more recent fMRI study of the UG (Tabibnia et al., 2008) found that increased activity in the ventrolateral PFC is correlated with increased acceptance of unfair offers, suggesting that this region may play the role originally attributed to the right DLPFC.

Other neuroeconomic studies have focused on how individuals track and respond to the moral status of others. Singer and colleagues (2004) examined neural responses to faces of people who had played either fairly (i.e. cooperatively) or unfairly in a sequential prisoner's dilemma game. Surprisingly, they found that faces of fair players, but not unfair players, elicited increased activity in the insula and the amygdala, regions widely, but not exclusively, associated with negative affect (Adolphs, 1999). In a second study, Singer and colleagues (2006) examined the interaction between (un)fair behavior and empathy. Both males and females exhibited signs of pain-empathy (increased activity in the fronto-insular cortex and ACC) when observing fair players receive a painful shock, but this effect was significantly reduced in males when the players receiving the shock had played unfairly. Males, moreover, exhibited increased reward-related activity in the nucleus accumbens (correlated with self-reported desire for revenge) when observing unfair players get shocked. In a similar vein, de Quervain and colleagues (2004) observed that reward-related activity in the caudate nucleus was associated with willingness to punish individuals who betrayed the subject's trust in a trust game. (A trust game is essentially a sequential prisoner's dilemma game in which cooperators must trust one another to continue cooperation.) A study by Delgado and colleagues (2005) examined the effect of reputation on moral-economic interaction. They had subjects play a trust game with fictional individuals who were characterized as good, bad, or neutral based on their personal histories. Their reputations affected subjects' willingness to trust them and modulated the level of activity in the caudate nucleus, partially overriding the effect of feedback during the game. King-Casas and colleagues (2006) used a trust game to examine the temporal dynamics of trust-development. They found that reward-related signals in the dorsal striatum were associated with the intention to trust and were shifted earlier in time as trust developed over the course of the game, mirroring effects observed in non-social reinforcement learning (Schultz et al., 1997). Taking a more molecular approach to the understanding of trust, Kosfeld and colleagues (2005) found that intranasal administration of oxytocin, a neuropeptide known for its role in social attachment and affiliation in non-human mammals (Insel and Young, 2001), increased trusting behavior.

Hsu and colleagues (2008) examined the neural bases of decisions concerning distributive justice, pitting deontological considerations for equality against utilitarian considerations in favor of maximizing aggregate benefits ("efficiency"). Their subjects allocated money to children in an orphanage, with some options favoring equality at the expense of efficiency and vice versa. Aversion to inequality was associated with increased activity in the insula, while activity in the putamen was positively correlated with the efficiency of the outcome. The caudate nucleus, in contrast, was sensitive to both factors, reflecting the subjective utility of the option. While at odds with the relatively simple dual-process theory presented above, these results are consistent with the Humean (1739/1978) conjecture (Greene et al., 2004; Greene, 2007b) that both deontological and utilitarian considerations ultimately have affective bases, despite the latter's greater dependence on controlled cognitive processing.

Conclusion

People often speak of a "moral sense" or a "moral faculty" (Hauser, 2006), but there is no single system within the human brain that answers to this description.

Rather, moral judgment emerges from a complex interaction among multiple neural systems whose functions are typically not (and maybe not ever be) specific to moral judgment (Greene and Haidt, 2002). The bulk of the research discussed above rightly emphasizes the role of emotion, in all of its functional and anatomical variety. At the same time, it is clear that controlled cognitive processing plays an important role in moral judgment, particularly in supporting judgments that run counter to prepotent emotional responses.

Three positive trends emerge from the foregoing discussion: First, we have seen a shift away from purely stimulus-based studies in favor of studies that associate patterns of neural activity with behavior. Second, and relatedly, we have seen an increased reliance on behavioral data, both in neuroscientific research and complementary behavioral studies. Third we have developed more ecologically valid paradigms involving real decisions, while recognizing that more stylized, hypothetical decisions can, like the geneticist's fruit fly, teach us valuable lessons. With regard to this issue, it is worth noting that in modern democracies our most important decisions are made indirectly by voters whose individual choices have little bearing on outcomes, and are thus effectively hypothetical.

Our current neuroscientific understanding of moral judgment is rather crude, conceptualized at the level of gross anatomical brain regions and psychological processes familiar from introspection. But, for all our ignorance, the physical basis of moral judgment is no longer a complete mystery. We've not only identified brain regions that are "involved" in moral judgment, but have begun to carve the moral brain at its functional joints.

Notes

[1] We defined "personal" moral dilemmas/harms as those involving actions that are (a) likely to cause serious bodily harm, (b) to a particular person, where (c) this harm does not result from deflecting an existing threat onto a different party (Greene et al., 2001). The first two criteria respectively exclude minor harms and harms to indeterminate "statistical" individuals. The third criterion aims to capture a sense of "agency," distinguishing between harms that are "authored" rather than merely "edited" by the agent in question. Recent research suggests that the dilemmas originally classified as "personal" and "impersonal" may be fruitfully classified in other ways (Mikhail, 2000; Royzman and Baron, 2002; Cushman et al., 2006; Waldmann and Dieterich, 2007; Moore et al., 2008).

Acknowledgments: Many thanks to Joe Paxton and Liane Young for their helpful comments.

Adolphs R (1999) Social cognition and the human brain. Trends Cogn Sci 3:469-479.

Adolphs R (2003) Cognitive neuroscience of human social behaviour. Nat Rev Neurosci 4:165-178.

Amodio DM, Frith CD (2006) Meeting of minds: the medial frontal cortex and social cognition. Nat Rev Neurosci 7:268-277.

Anderson SW, Bechara A, Damasio H, Tranel D, Damasio AR (1999) Impairment of social and moral behavior related to early damage in human prefrontal cortex. Nat Neurosci 2:1032-1037.

Aquinas T (unknown/2006) Summa Theologiae: Cambridge University Press.

Axelrod R, Hamilton W (1981) The evolution of cooperation. Science 211:1390-1396.

Bartels D (2008) Principled moral sentiment and the flexibility of moral judgment and decision making. Cognition 108:381-417.

Bechara A, Tranel D, Damasio H, Damasio AR (1996) Failure to respond autonomically to anticipated future outcomes following damage to prefrontal cortex. Cereb Cortex 6:215-225.

Beer JS, Heerey EA, Keltner D, Scabini D, Knight RT (2003) The Regulatory Function of Self-Conscious Emotion: Insights from Patients with Orbitofrontal Damage. Journal of Personality and Social Psychology 85:594-604.

Berkowitz L (1993) Aggression: its causes, consequences and control. Philadelphia, PA: Temple University Press.

Berthoz S, Armony JL, Blair RJ, Dolan RJ (2002) An fMRI study of intentional and unintentional (embarrassing) violations of social norms. Brain 125:1696-1708.

Berthoz S, Grezes J, Armony JL, Passingham RE, Dolan RJ (2006) Affective response to one's own moral violations. Neuroimage 31:945-950.

Blair RJ (1995) A cognitive developmental approach to mortality: investigating the psychopath. Cognition 57:1-29.

Blair RJ (2001) Neurocognitive models of aggression, the antisocial personality disorders, and psychopathy. J Neurol Neurosurg Psychiatry 71:727-731.

Blair RJ (2004) The roles of orbital frontal cortex in the modulation of antisocial behavior. Brain Cogn 55:198-208.

Blair RJ (2007) The amygdala and ventromedial prefrontal cortex in morality and psychopathy. Trends Cogn Sci 11:387-392.

Blair RJ, Jones L, Clark F, Smith M (1997) The psychopathic individual: a lack of responsiveness to distress cues? Psychophysiology 34:192-198.

Blonigen DM, Hicks BM, Krueger RF, Patrick CJ, Iacono WG (2005) Psychopathic personality traits: heritability and genetic overlap with internalizing and externalizing psychopathology. Psychol Med 35:637-648.

Botvinick MM, Braver TS, Barch DM, Carter CS, Cohen JD (2001) Conflict monitoring and cognitive control. Psychol Rev 108:624-652.

Buckner RL, Carrol DC (2007) Self-projection and the brain. Trends Cogn Sci 11:49-57.

Cacioppo J, Petty R, Kao C (1984) The efficient assessment of need for cognition. Journal of Personality Assessment 48:306-307.

Calder AJ, Lawrence AD, Young AW (2001) Neuropsychology of fear and loathing. Nat Rev Neurosci 2:352-363.

Chaiken S, Trope Y, eds (1999) Dual-Process Theories in Social Psychology. New York: Guilford Press.

Ciaramelli E, Muccioli M, Ladavas E, di Pellegrino G (2007) Selective deficit in personal moral judgment following damage to ventromedial prefrontal cortex. Social Cognitive and Affective Neuroscience 2:84-92.

Cohen JD, Perlstein WM, Braver TS, Nystrom LE, Noll DC, Jonides J, Smith EE (1997) Temporal dynamics of brain activation during a working memory task. Nature 386:604-608.

Colby A, Kohlberg L (1987) The measurement of moral judgment. Cambridge ; New York: Cambridge University Press.

Critchley HD, Elliott R, Mathias CJ, Dolan RJ (2000) Neural activity relating to generation and representation of galvanic skin conductance responses: a functional magnetic resonance imaging study. J Neurosci 20:3033-3040.

Crockett MJ, Clark L, Tabibnia G, Lieberman MD, Robbins TW (2008) Serotonin modulates behavioral reactions to unfairness. Science 320:1739.

Cushman F, Young L, Hauser M (2006) The role of conscious reasoning and intuition in moral judgment: testing three principles of harm. Psychol Sci 17:1082-1089.

Damasio AR (1994) Descartes' error : emotion, reason, and the human brain. New York: G.P. Putnam.

Darwin C (1871/2004) The Descent of Man. New York: Penguin.

Davis M, Whalen PJ (2001) The amygdala: vigilance and emotion. Mol Psychiatry 6:13-34.

de Quervain DJ, Fischbacher U, Treyer V, Schellhammer M, Schnyder U, Buck A, Fehr E (2004) The neural basis of altruistic punishment. Science 305:1254-1258.

Delgado MR, Frank R, Phelps EA (2005) Perceptions of moral character modulate the neural systems of reward during the trust game. Nat Neurosci 8:1611-1618.

Diergaarde L, Gerrits MA, Brouwers JP, van Ree JM (2005) Early amygdala damage disrupts performance on medial prefrontal cortex-related tasks but spares spatial learning and memory in the rat. Neuroscience 130:581-590.

Epstein S, Pacini R, DenesRaj V, Heier H (1996) Individual differences in intuitive-experiential and analytical-rational thinking styles. Journal of Personality and Social Psychology 71:390-405.

Finger EC, Marsh AA, Kamel N, Mitchell DG, Blair JR (2006) Caught in the act: the impact of audience on the neural response to morally and socially inappropriate behavior. Neuroimage 33:414-421.

Fischer JM, Ravizza M, eds (1992) Ethics: Problems and Principles. Fort Worth, TX: Harcourt Brace Jovanovich College Publishers.

Foot P (1978) The problem of abortion and the doctrine of double effect. In: Virtues and Vices. Oxford: Blackwell.

Fox MD, Snyder AZ, Vincent JL, Corbetta M, Van Essen DC, Raichle ME (2005) The human brain is intrinsically orgainzed into dynamic, anticorrelated functional networks. Proc Natl Acad Sci U S A 102:9673-9678.

Frederick S (2005) Cognitive Reflection and Decision Making. Journal of Economic Perspecitves 19:25-42.

Frith U (2001) Mind blindness and the brain in autism. Neuron 32:969-979.

Glenn A, Raine A, Schug R (2008) The neural correlates of moral decision-making in psychopathy. In.

Grattan LM, Eslinger PJ (1992) Long-term psychological consequences of childhood frontal lobe lesion in patient DT. Brain Cogn 20:185-195.

Greene J (in press) Social neuroscience and the soul's last stand. In: Social Neuroscience: Toward Understanding the Underpinnings of the Social Mind (Todorov A, Fiske S, Prentice D, eds). New York: Oxford University Press.

Greene J, Haidt J (2002) How (and where) does moral judgment work? Trends Cogn Sci 6:517-523.

Greene J, Morelli S, Lowenberg K, Nystrom L, Cohen J (2008a) Cognitive load selectively interferes with utilitarian moral judgment. Cognition 107:1144-1154.

Greene J, Lowenberg K, Paxton J, Nystrom L, Darley J, Cohen JD (2008b) Duty vs. the greater good: dissociable neural bases of deontological and utilitarian moral judgment in the context of keeping and breaking promises. In.

Greene J, Lindsell D, Clarke A, Lowenberg K, Nystrom L, Cohen J (2008c) Pushing moral buttons: The interaction between personal force and intention in moral judgment. In.

Greene JD (2007a) Why are VMPFC patients more utilitarian? A dual-process theory of moral judgment explains. Trends Cogn Sci 11:322-323; author reply 323-324.

Greene JD (2007b) The Secret Joke of Kant's Soul. In: Moral Psychology, Vol. 3: The Neuroscience of Morality: Emotion, Disease, and Development (Sinnott-Armstrong W, ed). Cambridge, MA: MIT Press.

Greene JD, Sommerville RB, Nystrom LE, Darley JM, Cohen JD (2001) An fMRI investigation of emotional engagement in moral judgment. Science 293:2105-2108.

Greene JD, Nystrom LE, Engell AD, Darley JM, Cohen JD (2004) The neural bases of cognitive conflict and control in moral judgment. Neuron 44:389-400.

Greene JD, Lowenberg K, Paxton J, Nystrom LE, Cohen JD (2008d) Neural dissociation between affective and cognitive moral disapproval. In.

Gusnard DA, Raichle ME (2001) Searching for a baseline: functional imaging and the resting human brain. Nat Rev Neurosci 2:685-694.

Gusnard DA, Akbudak E, Shulman GL, Raichle ME (2001) Medial prefrontal cortex and self-referential mental activity: relation to a default mode of brain function. Proc Natl Acad Sci U S A 98:4259-4264.

Guth W, Schmiittberger R, Schwarze B (1982) An experimental analysis of ultimatum bargaining. Journal of Economic Behavor and Organization 3:367-388.

Haidt J (2001) The emotional dog and its rational tail: A social intuitionist approach to moral judgment. Psychological Review 108:814-834.

Hamlin J, Wynn K, Bloom P (2007) Social evaluation by preverbal infants. Nature 450:557-560.

Hardman D (2008) Moral dilemmas: Who makes utilitarian choices. In.

Hare R (2003) Hare Psychopathy Checklist--Revised (PCL-R): 2nd Edition. In. Toronto: Multi-Health Systems, Inc.

Hare RD (1991) The Hare psychopathy checklist-revised. Toronto: Multi-Health Systems.

Harenski CL, Hamann S (2006) Neural correlates of regulating negative emotions related to moral violations. Neuroimage 30:313-324.

Hauser M (2006) The liver and the moral organ. Social Cognitive and Affective Neuroscience 1:214-220.

Heekeren HR, Wartenburger I, Schmidt H, Schwintowski HP, Villringer A (2003) An fMRI study of simple ethical decision-making. Neuroreport 14:1215-1219.

Heekeren HR, Wartenburger I, Schmidt H, Prehn K, Schwintowski HP, Villringer A (2005) Influence of bodily harm on neural correlates of semantic and moral decision-making. Neuroimage 24:887-897.

Hsu M, Anen C, Quartz SR (2008) The right and the good: distributive justice and neural encoding of equity and efficiency. Science 320:1092-1095.

Hume D (1739/1978) A treatise of human nature. In, 2d Edition (Selby-Bigge LA, Nidditch PH, eds), pp xix, 743. Oxford: Oxford University Press.

Insel TR, Young LJ (2001) The neurobiology of attachment. Nat Rev Neurosci 2:129-136.

Kahneman D (2003) A perspective on judgment and choice: mapping bounded rationality. Am Psychol 58:697-720.

Kant I (1785/1959) Foundation of the metaphysics of morals. Indianapolis: Bobbs-Merrill.

Kedia G, Berthoz S, Wessa M, Hilton D, Martinot J (in press) An agent harms a victim: a functional magnetic imaging study of specific moral emotions. Journal of Cognitive Neuroscience.

Kelley WM, Macrae CN, Wyland CL, Caglar S, Inati S, Heatherton TF (2002) Finding the self? An event-related fMRI study. J Cogn Neurosci 14:785-794.

Kiehl KA (2006) A cognitive neuroscience perspective on psychopathy: evidence for paralimbic system dysfunction. Psychiatry Res 142:107-128.

Kiehl KA, Smith AM, Hare RD, Mendrek A, Forster BB, Brink J, Liddle PF (2001) Limbic abnormalities in affective processing by criminal psychopaths as revealed by functional magnetic resonance imaging. Biol Psychiatry 50:677-684.

Killgore WD, Killgore DB, Day LM, Li C, Kamimori GH, Balkin TJ (2007) The effects of 53 hours of sleep deprivation on moral judgment. Sleep 30:345-352.

King JA, Blair RJ, Mitchell DG, Dolan RJ, Burgess N (2006) Doing the right thing: a common neural circuit for appropriate violent or compassionate behavior. Neuroimage 30:1069-1076.

Knoch D, Pascual-Leone A, Meyer K, Treyer V, Fehr E (2006) Diminishing reciprocal fairness by disrupting the right prefrontal cortex. Science 314:829-832.

Koenigs M, Tranel D (2007) Irrational economic decision-making after ventromedial prefrontal damage: evidence from the Ultimatum Game. J Neurosci 27:951-956.

Koenigs M, Young L, Adolphs R, Tranel D, Cushman F, Hauser M, Damasio A (2007) Damage to the prefrontal cortex increases utilitarian moral judgements. Nature 446:908-911.

Kohlberg L (1969) Stage and sequence: The cognitive-developmental approach to socialization. In: Handbook of socialization theory and research (Goslin DA, ed), pp 347-480. Chicago: Rand McNally.

Kosfeld M, Heinrichs M, Zak PJ, Fischbacher U, Fehr E (2005) Oxytocin increases trust in humans. Nature 435:673-676.

Leslie A, Mallon R, DiCorcia J (2006) Transgressors, victims, and cry babies: Is basic moral judgment spared in autism? Social Neuroscience 1:270-283.

Lieberman MD, Gaunt R, Gilbert DT, Trope Y (2002) Reflection and reflexion: A social cognitive neuroscience approach to attributional inference. Advances in Experimental Social Psychology 34:199-249.

Luo Q, Nakic M, Wheatley T, Richell R, Martin A, Blair RJ (2006) The neural basis of implicit moral attitude--an IAT study using event-related fMRI. Neuroimage 30:1449-1457.

Macmillan M (2000) An odd kind of fame. Cambridge, MA: MIT Press.

Maddock RJ (1999) The retrosplenial cortex and emotion: new insights from functional neuroimaging of the human brain. Trends Neurosci 22:310-316.

Mason MF, Norton MI, Van Horn JD, Wegner DM, Grafton ST, Macrae CN (2007) Wandering minds: the default network and stimulus-independent thought. Science 315:393-395.

Mendez MF, Anderson E, Shapira JS (2005) An investigation of moral judgement in frontotemporal dementia. Cogn Behav Neurol 18:193-197.

Mikhail J (2000) Rawls' Linguistic Analogy: A Study of the "Generative Grammar" Model of Moral Theory Described by John Rawls in *A Theory of Justice*. In: Cornell University.

Mill JS (1861/1998) Utilitarianism. In: (Crisp R, ed). New York: Oxford University Press.

Miller EK, Cohen JD (2001) An integrative theory of prefrontal cortex function. Annu Rev Neurosci 24:167-202.

Moll J, de Oliveira-Souza R (2007) Moral judgments, emotions and the utilitarian brain. Trends Cogn Sci 11:319-321.

Moll J, Eslinger PJ, Oliveira-Souza R (2001) Frontopolar and anterior temporal cortex activation in a moral judgment task: preliminary functional MRI results in normal subjects. Arq Neuropsiquiatr 59:657-664.

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

Moll J, Krueger F, Zahn R, Pardini M, de Oliveira-Souza R, Grafman J (2006) Human fronto-mesolimbic networks guide decisions about charitable donation. Proc Natl Acad Sci U S A 103:15623-15628.

Moll J, de Oliveira-Souza R, Eslinger PJ, Bramati IE, Mourao-Miranda J, Andreiuolo PA, Pessoa L (2002b) The neural correlates of moral sensitivity: a functional magnetic resonance imaging investigation of basic and moral emotions. J Neurosci 22:2730-2736.

Moll J, de Oliveira-Souza R, Moll F, Ignacio F, Bramati I, Caparelli-Daquer E, Eslinger P (2005) The moral affiliations of disgust: a functional MRI study. Cognitive and behavioral neurology 18:68-78.

Moore A, Clark B, Kane M (2008) Who shalt not kill?: Individual differences in working memory capacity, executive control, and moral judgment. Psychological Science 19:549-557.

Muller JL, Sommer M, Dohnel K, Weber T, Schmidt-Wilcke T, Hajak G (2008) Disturbed prefrontal and temporal brain function during emotion and cognition interaction in criminal psychopathy. Behav Sci Law 26:131-150.

Oxford M, Cavell TA, Hughes JN (2003) Callous/unemotional traits moderate the relation between ineffective parenting and child externalizing problems: a partial replication and extension. J Clin Child Adolesc Psychol 32:577-585.

Petrinovich L, O'Neill P, Jorgensen M (1993) An empirical study of moral intuitions: Toward an evolutionary ethics. Journal of Personality and Social Psychology 64:467-478.

Phan KL, Wager T, Taylor SF, Liberzon I (2002) Functional Neuroanatomy of Emotion: A Meta-Analysis of Emotion Activation Studies in PET and fMRI. Neuroimage 16:331-348.

Phillips ML, Young AW, Senior C, Brammer M, Andrew C, Calder AJ, Bullmore ET, Perrett DI, Rowland D, Williams SC, Gray JA, David AS (1997) A specific neural substrate for perceiving facial expressions of disgust. Nature 389:495-498.

Piaget J (1965) The moral judgement of the child. In. New York: Free Press.

Posner MI, Snyder CRR (1975) Attention and cognitive control. In: Information processing and cognition (Solso RL, ed), pp 55-85. Hillsdale, NJ: Erlbaum.

Raine A, Yang Y (2006) Neural foundations to moral reasoning and antisocial behavior. Social Cognitive and Affective Neuroscience 1:203-213.

Raine A, Lencz T, Bihrle S, LaCasse L, Colletti P (2000) Reduced prefrontal gray matter volume and reduced autonomic activity in antisocial personality disorder. Arch Gen Psychiatry 57:119-127; discussion 128-119.

Raine A, Buchsbaum MS, Stanley J, Lottenberg S, Abel L, Stoddard J (1994) Selective reductions in prefrontal glucose metabolism in murderers. Biol Psychiatry 36:365-373.

Rilling J, Glenn A, Jairam M, Pagnoni G, Goldsmith D, Elfenbein H, Lilienfeld S (2007) Neural correlates of social cooperation and non-cooperation as a function of psychopathy. Biol Psychiatry 61:1260-1271.

Robertson D, Snarey J, Ousley O, Harenski K, DuBois Bowman F, Gilkey R, Kilts C (2007) The neural processing of moral sensitivity to issues of justice and care. Neuropsychologia 45:755-766.

Royzman EB, Baron J (2002) The preference for indirect harm. Social Justice Research 15:165-184.

Rozin P, Lowery L, Ebert R (1994) Varieties of disgust faces and the structure of disgust. Journal of Personality and Social Psychology 66:870-881.

Rozin P, Lowery L, Imada S, Haidt J (1999) The CAD triad hypothesis: a mapping between three moral emotions (contempt, anger, disgust) and three moral codes (community, autonomy, divinity). J Pers Soc Psychol 76:574-586.

Sanfey AG, Rilling JK, Aronson JA, Nystrom LE, Cohen JD (2003) The neural basis of economic decision-making in the Ultimatum Game. Science 300:1755-1758.

Saver J, Damasio A (1991) Preserved access and processing of social knowledge in a patient with acquired sociopathy due to ventromedial frontal damage. Neuropsychologia 29:1241-1249.

Saxe R, Carey S, Kanwisher N (2004) Understanding other minds: Linking Developmental Psychology and Functional Neuroimaging. Annu Rev Psychol 55:87-124.

Schacter DL, Addis DR, Buckner RL (2007) Remembering the past to imagine the future: the prospective brain. Nat Rev Neurosci 8:657-661.

Schaich Borg J, Lieberman D, Kiehl KA (2008) Infection, Incest, and Iniquity: Investigating the neural correlates of disgust and morality. Journal of Cognitive Neuroscience 20:1-19.

Schaich Borg J, Hynes C, Van Horn J, Grafton S, Sinnott-Armstrong W (2006) Consequences, Action, and Intention as Factors in Moral Judgments: An fMRI Investigation. Journal of Cognitive Neuroscience 18:803-817.

Scherer K, Schorr A, Johnstone T, eds (2001) Appraisal Processes in Emotion. New York: Oxford Univeristy Press.

Schneider F, Habel U, Kessler C, Posse S, Grodd W, Muller-Gartner HW (2000) Functional imaging of conditioned aversive emotional responses in antisocial personality disorder. Neuropsychobiology 42:192-201.

Schoenbaum G, Roesch M (2005) Orbitofrontal cortex, associative learning, and expectancies. Neuron 47.

Schultz W, Dayan P, Montague PR (1997) A neural substrate of prediction and reward. Science 275:1593-1599.

Shin L, Dougherty D, Orr S, Pitman R, Lasko M, Macklin M, Alpert N, Fischman A, Rauch S (2000) Activation of anterior paralimbic structures during guilt-related script-driven imagery. Biol Psychiatry 48:43-50.

Singer T, Kiebel S, Winston J, Dolan R, Frith C (2004) Brain response to the acquired moral status of faces. Neuron 41:653-662.

Singer T, Seymour B, O'Doherty JP, Stephan KE, Dolan RJ, Frith CD (2006) Empathic neural responses are modulated by the perceived fairness of others. Nature 439:466-469.

Smith EE, Jonides J (1997) Working memory: a view from neuroimaging. Cognit Psychol 33:5-42.

Tabibnia G, Satpute AB, Lieberman MD (2008) The sunny side of fairness: preference for fairness activates reward circuitry (and disregarding unfairness activates self-control circuitry). Psychol Sci 19:339-347.

Takahashi H, Yahata N, Koeda M, Matsuda T, Asai K, Okubo Y (2004) Brain activation associated with evaluative processes of guilt and embarrassment: an fMRI study. Neuroimage 23:967-974.

Thomson J (1985) The Trolley Problem. Yale Law Journal 94:1395-1415.

Turiel E (2006) Thought, emotions and social interactional processes in moral development. In: Handbook of Moral Develpment (Killen M, Smetana J, eds). Mahwah, NJ: Lawrence Erlbaum Assoc.

Valdesolo P, DeSteno D (2006) Manipulations of emotional context shape moral judgment. Psychol Sci 17:476-477.

Viding E, Blair RJ, Moffitt TE, Plomin R (2005) Evidence for substantial genetic risk for psychopathy in 7-year-olds. J Child Psychol Psychiatry 46:592-597.

Vollm B, Richardson P, Stirling J (2004) Neurobiological substrates of antisocial and borderline personality disorders: prelimnary results from an fMRI study. Criminal Behavior and Mental Health 14:39-54.

Waal d (2006) Primates and Philosophers: How Morality Evolved. Princeton, NJ: Princeton University Press.

Waldmann MR, Dieterich JH (2007) Throwing a bomb on a person versus throwing a person on a bomb: intervention myopia in moral intuitions. Psychol Sci 18:247-253.

Warneken F, Hare B, Melis AP, Hanus D, Tomasello M (2007) Spontaneous Altruism by Chimpanzees and Young Children. PLoS Biol 5:e184.

Wheatley T, Haidt J (2005) Hypnotic disgust makes moral judgments more severe. Psychol Sci 16:780-784.

Yang Y, Raine A, Narr K, Lencz T, Toga A (2006) Amygdala volume reduction in psychopaths. In: Socieity for Research in Psychopathology.

Yang Y, Raine A, Lencz T, Bihrle S, LaCasse L, Colletti P (2005) Volume reduction in prefrontal gray matter in unsuccessful criminal psychopaths. Biol Psychiatry 57:1103-1108.

Young L, Saxe R (2008) The neural basis of belief encoding and integration in moral judgment. Neuroimage 40:1912-1920.

Young L, Saxe R (in press) An fMRI investigation of spontaneous mental state inference for moral judgment. Journal of Cognitive Neuroscience.

Young L, Cushman F, Hauser M, Saxe R (2007) The neural basis of the interaction between theory of mind and moral judgment. Proc Natl Acad Sci U S A.

Young L, Camprodon J, Hauser M, Pascual-Leone A, Saxe R (2008) Disrupting neural mechanisms involved in belief attribution impairs moral judgment. In.

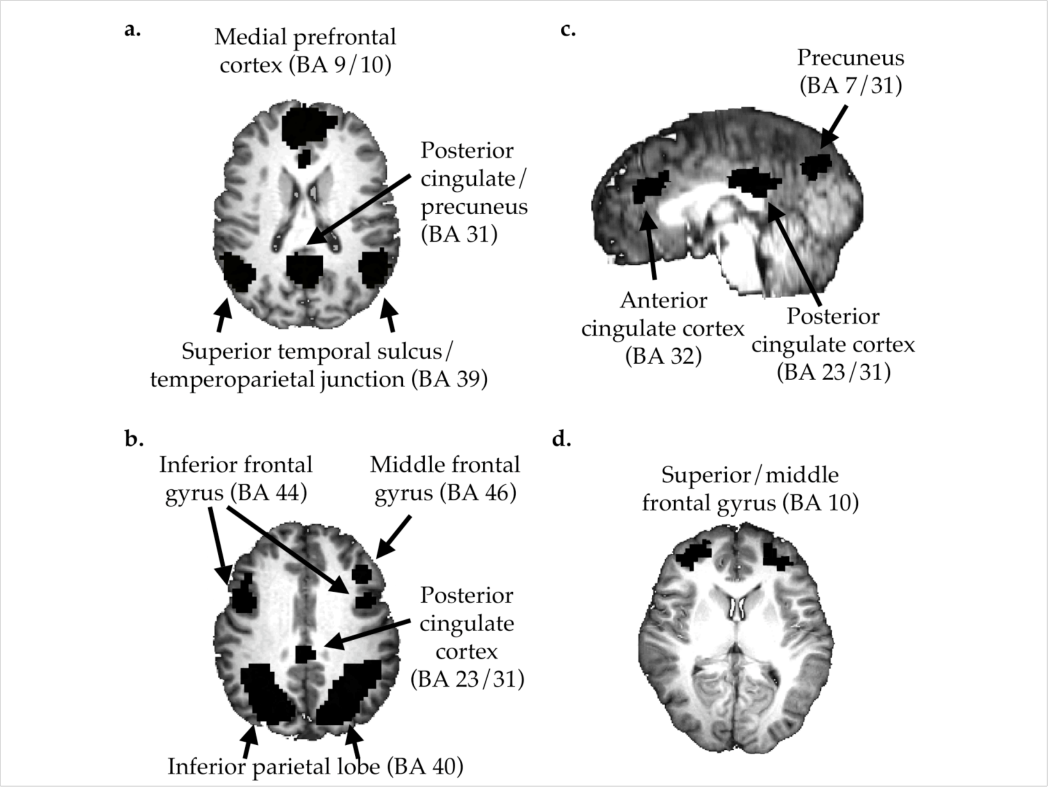

a.
Medial prefrontal cortex (BA 9/10)
Posterior cingulate/ precuneus (BA 31)
Superior temporal sulcus/ temperoparietal junction (BA 39)
c.
Precuneus (BA 7/31)
Anterior cingulate cortex (BA 32)
Posterior cingulate cortex (BA 23/31)
b.
Inferior frontal gyrus (BA 44)
Middle frontal gyrus (BA 46)
Posterior cingulate cortex (BA 23/31)
Inferior parietal lobe (BA 40)
d.
Superior/middle frontal gyrus (BA 10)

Fig 1

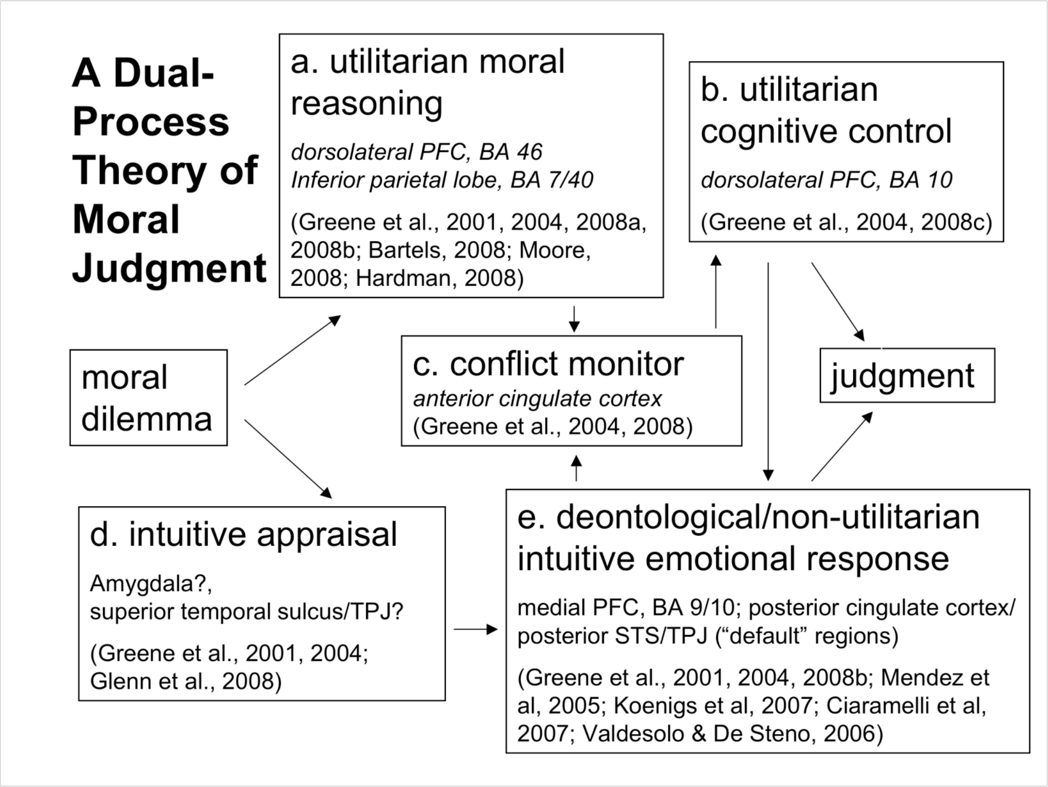
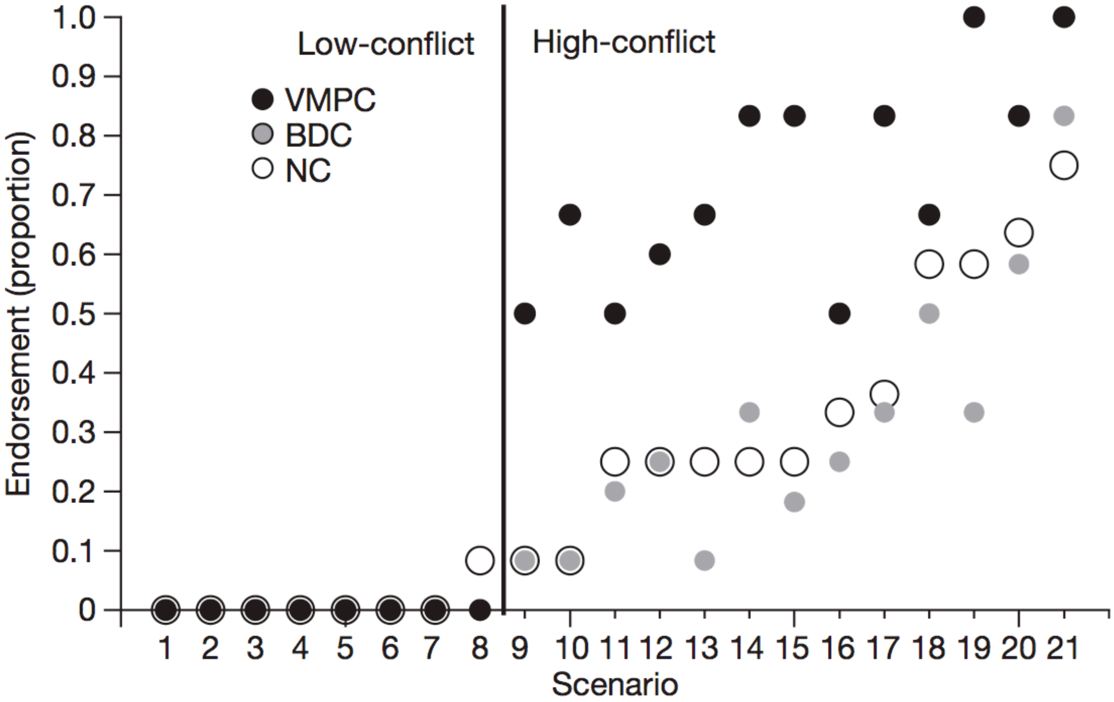

A Dual-Process Theory of Moral Judgment
a. utilitarian moral reasoning
dorsolateral PFC, BA 46
Inferior parietal lobe, BA 7/40
(Greene et al., 2001, 2004, 2008a, 2008b; Bartels, 2008; Moore, 2008; Hardman, 2008)
b. utilitarian cognitive control
dorsolateral PFC, BA 10
(Greene et al., 2004, 2008c)
moral dilemma
c. conflict monitor
anterior cingulate cortex
(Greene et al., 2004, 2008)
judgment
d. intuitive appraisal
Amygdala?,
superior temporal sulcus/TPJ?
(Greene et al., 2001, 2004; Glenn et al., 2008)
e. deontological/non-utilitarian intuitive emotional response
medial PFC, BA 9/10; posterior cingulate cortex/ posterior STS/TPJ ("default" regions)
(Greene et al., 2001, 2004, 2008b; Mendez et al, 2005; Koenigs et al, 2007; Ciaramelli et al, 2007; Valdesolo & De Steno, 2006)

Fig 2



Low-conflict
High-conflict
● VMPC
● BDC
○ NC
Endorsement (proportion)
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
Scenario

Fig 3