

Published in final edited form as:

*Neuroimage*. 2010 February 1; 49(3): 2707–2716. doi:10.1016/j.neuroimage.2009.10.062.

## A functional imaging investigation of moral deliberation and moral intuition

Carla L. Harenski<sup>a,\*</sup>, Olga Antonenko<sup>c</sup>, Matthew S. Shane<sup>a</sup>, and Kent A. Kiehl<sup>a,b</sup>

<sup>a</sup>The Mind Research Network, 1101 Yale Boulevard Northeast, Albuquerque, NM 87131, USA

<sup>b</sup>Departments of Psychology and Neuroscience, University of New Mexico, Albuquerque, NM, USA

<sup>c</sup>University of California, Berkeley, Berkeley, CA, USA

### Abstract

Prior functional imaging studies of moral processing have utilized ‘explicit’ moral tasks that involve moral deliberation (e.g., reading statements such as ‘he shot the victim’ and rating the moral appropriateness of the behavior) or ‘implicit’ moral tasks that involve moral intuition (e.g., reading similar statements and memorizing them for a test but not rating their moral appropriateness). Although the neural mechanisms underlying moral deliberation and moral intuition may differ, these have not been directly compared. Studies using explicit moral tasks have reported increased activity in several regions, most consistently the medial prefrontal cortex and temporo-parietal junction. In the few studies that have utilized implicit moral tasks, medial prefrontal activity has been less consistent, suggesting the medial prefrontal cortex is more critical for moral deliberation than moral intuition. Thus, we hypothesized that medial prefrontal activity would be increased during an explicit, but not an implicit, moral task. Participants ( $n = 28$ ) were scanned using fMRI while viewing 50 unpleasant pictures, half of which depicted moral violations. Half of the participants rated pictures on moral violation severity (explicit task) while the other half indicated whether pictures occurred indoors or outdoors (implicit task). As predicted, participants performing the explicit, but not the implicit, task showed increased ventromedial prefrontal activity while viewing moral pictures. Both groups showed increased temporo-parietal junction activity while viewing moral pictures. These results suggest that the ventromedial prefrontal cortex may contribute more to moral deliberation than moral intuition, whereas the temporo-parietal junction may contribute more to moral intuition than moral deliberation.

### Introduction

Moral judgment is a complex process involving a combination of automatic intuitions and deliberate reasoning, which contribute to an individual’s personal moral values, i.e., what

© 2009 Elsevier Inc. All rights reserved.

\*Corresponding author. Fax: +1 505 272 8002. charenski@mrn.org (C.L. Harenski).

#### Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.neuroimage.2009.10.062.

one believes to be ‘right’ and ‘wrong.’ To identify the neural mechanisms underlying these processes, functional imaging studies have explored a wide variety of moral processing, ranging from passive viewing of pictures depicting moral violations to the evaluation of complex moral dilemmas (for reviews see Greene and Haidt, 2002; Moll et al., 2005). These studies have identified a consistent set of brain regions implicated in the evaluation of moral stimuli, including the medial prefrontal cortex, posterior temporal cortex including the temporo-parietal junction and superior temporal sulcus, and, less consistently, posterior cingulate/precuneus and anterior temporal cortex. The roles of these and other brain regions in moral judgment are becoming increasingly understood, especially in light of several studies that have directly compared different types of moral processing (Greene et al., 2004; Heekeren et al., 2005; Young et al., 2007).

The majority of these studies have utilized ‘explicit’ moral tasks, tasks in which participants are presented with morally salient stimuli and instructed to rate their moral appropriateness. We refer to these tasks as ‘explicit’ because participants are made fully aware that the task involves moral content which they will be required to evaluate. The moral stimuli range from simple statements to complex dilemmas, often describing a protagonist who commits a moral violation such as stealing or physically harming another individual (Greene et al., 2001, 2004; Moll et al., 2002a; Heekeren et al., 2003, 2005; Schaich Borg et al., 2006; Robertson et al., 2007; Young et al., 2007; Prehn et al., 2008; Young and Saxe, 2008). These are contrasted with similar statements or dilemmas that do not contain moral violations. In these studies neural activity in response to moral stimuli can be inferred to represent moral deliberation, which includes the process of evaluating stimuli on moral appropriateness, and the identification of a morally inappropriate act. This type of task is different from an ‘implicit’ moral task, in which participants are presented with morally salient stimuli but are not instructed to evaluate their moral appropriateness, nor are they even informed that a distinction between morally neutral or morally inappropriate actions characterizes the stimuli (Moll et al., 2002b; Harenski and Hamann, 2006; Moll et al., 2007; Schaich Borg et al., 2008). In these studies neural activity in response to moral stimuli is less likely to represent moral deliberation, and can be inferred to represent moral intuition. Moral intuitions may include spontaneous, unsolicited attention directed towards cues that have potential moral salience such as a person in distress, weapons or other objects (pictorial stimuli), or emotion-laden words such as ‘assault’ or ‘betrayed’ (verbal stimuli).

Several theoretical perspectives on moral judgment have highlighted the role of moral intuition and moral deliberation (for a review see Hauser, 2006), though the manner in which these processes contribute to moral judgment and their relative importance (Haidt, 2001; Pizarro and Bloom, 2003) has been debated. Different perspectives have been offered regarding the nature of moral intuitions that are automatically elicited by the perception of a morally salient event (e.g., emotion-based (Haidt, 2001) or cause/intention-based (Hauser, 2006)), and whether moral deliberation occurs prior to (Greene et al., 2004) or following (Haidt, 2001; Hauser, 2006) moral judgment. These differing perspectives emphasize the complexity of the moral judgment process, and raise intriguing questions regarding which brain regions implicated in moral processing contribute to moral intuition and/or deliberation. However, since prior studies have mostly used explicit moral tasks, which focus on moral deliberation, the neural correlates of moral intuition relative to moral

deliberation remain largely unknown. The importance of investigating moral intuition is underscored by the fact that in the ‘real world’ (outside of the MRI scanner), individuals do not typically evaluate their surroundings with the intention to identify morally salient information. In everyday encounters an individual may unexpectedly find oneself in a moral dilemma, or witness an action that may or may not constitute a moral violation. Whether morally salient events are identified as such depends on numerous factors such as the individuals’ own set of moral values, or those in accordance with a particular culture (Schweder et al., 1987; Haidt et al., 1993). Moreover, the presence of a moral violation may not always be obvious. Morally salient cues might be attended but not result in moral deliberation or moral judgment. Overall, the distinction between moral intuition and moral deliberation emphasizes the importance of identifying neural systems underlying both forms of moral processing.

There is some evidence that certain brain regions implicated in moral processing may differentially contribute to moral intuition and moral deliberation. In one study that examined moral intuition by using an implicit moral task (Harenski and Hamann, 2006), participants passively viewed moral and non-moral pictures without being instructed to evaluate their moral appropriateness. Both sets of pictures were matched for emotional and social content, but only one set depicted moral violations. Participants were not aware of the moral/non-moral picture distinction (confirmed by post-scan interviews) and thus did not engage in moral deliberation when viewing moral pictures. Despite this, increased activity during moral relative to non-moral picture viewing occurred in brain regions that have been implicated in prior studies of moral processing, including the temporo-parietal junction and posterior cingulate. This finding suggested that these regions are involved in moral intuition. In contrast, the medial prefrontal cortex, a brain region that has been consistently implicated in prior studies using explicit moral tasks, did not show increased activity during moral relative to non-moral picture viewing. Increased medial prefrontal activity in response to moral pictures occurred only when participants were instructed to decrease their emotional responses to the pictures, thus representing an interaction between moral intuitions and the intentional down-regulation of associated emotional responses. Another study that used an implicit moral task also found no medial prefrontal activity when moral statements were contrasted with non-moral statements of similar emotional valence and arousal (Schaich Borg et al., 2008). In contrast to these findings, Moll et al. (2007) reported increased medial prefrontal activity during passive reading of statements designed to evoke different types of ‘moral emotion,’ such as guilt and compassion. An earlier study by the same researchers also reported increased medial prefrontal activity during passive viewing of pictures depicting moral violations relative to those that did not (Moll et al., 2002a,b). However, the moral and non-moral pictures differed in some critical aspects other than moral content. For example, most of the moral pictures contained social scenes, whereas many of the non-moral pictures depicted objects. Given the role of the medial prefrontal cortex in social cognition (Amodio and Frith, 2006), it is difficult to know whether the ventromedial activity was due to moral processing per se or a broader factor associated with social cognition.

Thus, while the role of the medial prefrontal cortex in moral deliberation is well established, its role in moral intuition is less clear. It is possible that the medial prefrontal cortex contributes more to moral deliberation than moral intuition. This suggestion is consistent

with the demonstrated role of the medial prefrontal cortex in simple and complex decision making (Cunningham et al., 2003; Paulus and Frank, 2003; Sanfey et al., 2003; Fellows and Farah, 2005, 2007), particularly in conjunction with emotional responses (Bechara et al., 1997, 1999, 2000). The functions of other brain regions implicated in moral processing, including the temporo-parietal junction and posterior cingulate, have been attributed to theory of mind (Gallagher and Frith, 2003; Ruby and Decety, 2003; Saxe and Kanwisher, 2003; Ciaramidaro et al., 2007) and emotional and self-reflective processing (Fink et al., 1996; Maddock, 1999; Damasio et al., 2000; Vogt and Laureys, 2005; Johnson et al., 2006). In contrast to the medial prefrontal cortex, these regions might contribute more to moral intuition than moral deliberation, or contribute to both processes. However, these hypotheses are currently tentative given that few studies have explored moral intuitive processing in the absence of moral deliberation, and none have directly compared the two processes.

To test these hypotheses, the current study used functional magnetic resonance imaging (fMRI) to evaluate brain activity during the performance of two different moral processing tasks. In the first task, referred to as the 'explicit moral task,' participants viewed unpleasant pictures that did or did not contain moral violations (e.g., a hand breaking into a house vs. a mutilated hand), as well as neutral pictures (e.g., a hand being fingerprinted), and rated each picture on the degree of moral violation severity present in the picture (Harenski et al., 2008). A second group of participants was recruited to perform an 'implicit moral task.' In this task, participants viewed the same pictures as the participants who performed the explicit moral task, but judged whether each picture occurred indoors or outdoors and were thus not made aware of the moral/non-moral picture distinction. The between-group design was chosen over a within-group design to avoid carryover effects from the explicit moral task to the implicit moral task. The implicit task was designed to ensure that participants did not engage in moral deliberation during the task by not informing these participants that the study involved moral processing (this approach has been successful in our prior work; see Harenski and Hamann, 2006). In a within-group design, the implicit task would always have to be presented before the explicit task, to reduce the possibility of spontaneous moral deliberation occurring during the implicit task. This would also require re-presenting the stimuli during the explicit task, increasing the risk of explicit task-specific habituation effects. Thus, the between-group design ensured that all effects of each task type were independent of the other.

The hypothesis was that medial prefrontal activity would be significantly increased during moral picture processing in participants who performed the explicit task, but not those who performed the implicit task. In contrast, activity in the temporo-parietal junction and posterior cingulate, which has previously been reported in implicit moral processing tasks that did not involve moral deliberation (Harenski and Hamann, 2006), was expected during moral picture processing in both explicit and implicit task participants (since moral intuitions were expected to occur during both tasks). Whether these and/or other brain regions involved in moral processing would show increased responses during the implicit relative to explicit moral task condition was an open question.

## Methods

### Participants

Thirty healthy, right handed female adults (age range 18–34 years) were recruited from Hartford Hospital (Hartford, CT) and a local liberal arts college (Trinity College, Hartford, CT) via advertisements and word of mouth. Sixteen of these participants performed the ‘explicit moral task,’ and were also included in a prior study that examined gender differences in explicit moral processing (Harenski et al., 2008). The other 14 participants were recruited for the present study, and performed the ‘implicit moral task.’ Both participant groups were matched on age (mean explicit task=23.9 (SD=3.85); mean implicit task=24.5 (SD=3.94);  $p=.70$ ) and of similar education level. Two participants from the explicit task group were excluded: one due to excessive head motion during scanning (>5 mm), and another due to poor task performance (the participant missed several ratings during the scan). All participants provided written informed consent, and the study was conducted in accordance with institutional ethical standards.

### Stimuli and tasks

Three sets of pictures (25 moral, 25 non-moral, 25 neutral) were selected mostly from the International Affective Picture System (IAPS; Lang, Bradley, and Cuthbert, 1995), and supplemented with pictures from the popular media (examples shown in Fig. 1). Moral pictures were unpleasant social scenes depicting a moral violation (e.g., an abusive situation; a drunk driver). Non-moral pictures depicted unpleasant social scenes without moral content (e.g., an argument; an angry driver). Neutral pictures depicted affectively neutral social scenes without moral content (e.g., a conversation and a normal driver). Moral and non-moral pictures were a subset of those utilized in Harenski and Hamann (2006), and were matched on emotional arousal and social complexity.

Participants in the explicit moral task group were informed that they would see a series of different pictures depicting people and events. For each picture, they were instructed to determine whether it represented a moral violation (an action or attitude that the participant considered to be morally wrong) and to rate the severity of the moral violation on a scale from 1 (none) to 5 (severe). If the picture did not represent a moral violation, participants were instructed to give a rating of 1. Emphasis was placed on asking the participants to make ratings based on their own system of moral values, rather than what others or society would consider a moral violation. Participants in the implicit moral task group were presented with the identical picture set, but were instructed to determine whether each picture occurred indoors or outdoors. These participants were not informed of the moral/non-moral distinction across pictures.

Following the instructions, participants entered the scanner and completed five practice trials to ensure that they understood how to perform the task. Each trial proceeded as follows: In the explicit task, a picture which did or did not contain a moral violation was first displayed for 6 s. Next, a rating scale was shown. The rating scale was displayed in continuous presentation format, such that a red bar began at ‘1’ (none) and progressed to ‘5’ (severe) over a period of 4 s (see Fig. 1). The participant pressed a button to stop the bar

when it reached the violation severity rating that they wished to give. Following the rating, a 4-s rest period occurred during which a black screen with a white fixation cross was displayed. The implicit task was the same as the explicit task, with the following exceptions: during the 6-s picture presentation, participants determined whether the picture occurred indoors or outdoors, and during the 4-s rating period that followed each picture they were presented with a screen reading 'Indoor/Outdoor?' They were instructed to press one button with their index finger if the picture occurred indoors, and a different button with their middle finger if the picture occurred outdoors. Due to technical error, online ratings were not obtained for one explicit task participant.

Moral, non-moral, and neutral picture trials were presented in a randomized order, and were interspersed with 25 fixation trials of the same duration as picture trials. The fixation trials were randomized in the same manner as the other trial types, which provided an inherent jittering of the intertrial interval. The 100 total trials were presented across two separate runs. Images were rear-projected into the scanner using an LCD projector, controlled by a PC computer. Tasks were designed and presented and responses were recorded using Presentation software (Neurobehavioral Systems, Davis, CA).

Following scanning, participants viewed the picture set for a second time and rated them on degree of emotional arousal (1=low, 5=high). These ratings served as a manipulation check to verify that moral and non-moral pictures were considered similar in emotional arousal in the current study participants.

### **MRI data acquisition and analysis**

Whole-brain imaging data were obtained using a Siemens 3T Allegra MRI scanner. The gradient echo planar sequence (TR=1500 ms, TE=27 ms, FA=65, FOV 24×24 cm, 64×64 matrix, 3.44 by 3.44 mm in plane resolution, 5 mm slice thickness, 30 slices) effectively covered the entire brain [150 mm]. Within the ventral region of the medial prefrontal cortex, notable signal dropout was only present ventral to  $z = -16$  (see Supplementary Fig. 1). A total of 480 scans were obtained in each of 2 scan runs. Head movement was limited by padding and restraint. Functional images were motion corrected, normalized to a standard template, and spatially smoothed (8 FWHM) using SPM2 ([www.fil.ion.ucl.ac.uk/spm](http://www.fil.ion.ucl.ac.uk/spm)). Low frequency noise was removed using a high-pass filter (Holmes et al., 1997).

Individual participant data were analyzed using the general linear model with a random effects analysis in SPM2 (Friston et al., 1994). Picture presentations (moral, non-moral, neutral) and the rating period were modeled as separate events. The event of interest, picture presentation, was modeled with a 6-s duration and convolved with the standard hemodynamic response function. Statistical maps were computed for each of the picture conditions in each individual and each group. The linear contrast of moral vs. non-moral picture viewing assessed hemodynamic responses associated with processing moral content, while controlling for general effects of emotional arousal and social content. An independent samples t-test was used to directly compare brain activity associated with moral vs. non-moral picture viewing in participants performing the explicit (moral rating) task to those performing the implicit (indoor/outdoor) task. Parameter estimates of individual-participant activity in brain regions showing increased activity in the explicit vs. implicit groups were

extracted for each of the four conditions (explicit moral, explicit non-moral, implicit moral, implicit non-moral) from the activated clusters (i.e., averaged across all voxels) in each region using the Marsbar toolbox for SPM2 (<http://marsbar.sourceforge.net/>). To ensure that any observed differences were specifically due the effects of the moral condition, we conducted the following additional between-group comparisons: between-group comparison of moral vs. neutral picture viewing, between-group comparison of non-moral vs. neutral picture viewing, and between-group comparison of moral picture viewing (without reference to a control condition).

In a subsequent analysis, moral vs. non-moral picture viewing was again contrasted using a model in which the violation severity ratings (explicit task participants) and post-scan emotional arousal ratings (explicit and implicit task participants) were entered as covariates of no interest to determine whether they influenced the main effects of moral > non-moral picture viewing, and of interest to explore whether activity in any brain regions during picture viewing was correlated with the subsequent ratings.

The between-group design, which was used to avoid carryover effects between explicit and implicit moral processing, is typically less powerful than a within-group design. Thus, statistical maps for all between-group comparisons were thresholded at  $p < .005$  (uncorrected) with an extent threshold of 5 contiguous voxels. Statistical maps for within-group comparisons (e.g., moral > non-moral contrast in the participants who performed either the explicit or implicit task) were thresholded at  $p < .001$ , uncorrected. Small-volume correction (SVC) analyses were performed on activations within a priori regions of interest (medial prefrontal cortex, temporo-parietal junction, and posterior cingulate), with anatomical boundaries determined based on previous studies (Berthoz et al., 2006; Greene et al., 2001, 2004; Harenski and Hamann, 2006; Harenski et al., 2008; Heekeren et al., 2003, 2005; Schaich Borg et al., 2006, 2008; Moll et al., 2002a,b, 2005; Young et al., 2007; Young and Saxe, 2008). The local maxima within each region of interest was identified and used as the center coordinate of the region of interest with a sphere radius of 10 mm (anterior medial prefrontal cortex, BA 9/10; ventromedial prefrontal cortex/orbitofrontal cortex, BA 10/11), 14 mm (temporo-parietal junction, BA 39), and 12 mm (posterior cingulate, BA 31), which was corrected using a threshold of  $p < .05$  (see also Moll et al., 2007). Activations were overlaid on a representative high-resolution structural T1-weighted image from a single subject from the SPM2 canonical image set, coregistered to Montreal Neurological Institute (MNI) space. All coordinates are reported in MNI space.

## Results

### Behavioral results

**Task performance**—Violation severity ratings across each of the three picture conditions (moral, non-moral, neutral) made by participants performing the explicit task are reported in Harenski et al. (2008). Participants rated moral pictures higher on moral violation severity than non-moral pictures [ $F_{1,12} = 209.93, p < .0001$ ] or neutral pictures [ $F_{1,12} = 480.10, p < .0001$ ] (Fig. 2A). Participants performing the implicit task were more than 85% accurate in the identification of indoor/outdoor setting across all picture conditions. Accuracy was higher in the neutral condition relative to the moral [ $F_{1,13} = 53.30, p < .0001$ ] and non-moral

[ $F_{1,13} = 30.60, p < .0001$ ] conditions (Fig. 2B). No significant differences were present between the moral and non-moral conditions [ $F_{1,13} = 0.50, p = .49$ ].

We did not obtain violation severity ratings from participants who performed the implicit task. However, we did obtain these ratings from a different group of implicit task participants (passive viewing of moral and non-moral pictures) in a previous study (Harenski and Hamann, 2006). The violation severity ratings of those participants were similar to those of participants who performed the explicit task in the current study (moral: mean explicit = 3.98, mean implicit = 3.90,  $p = .71$ ; non-moral: mean explicit = 1.77, mean implicit = 1.95,  $p = .45$ ).

**Post-scan ratings**—Since moral and non-moral pictures were matched *a priori* on emotional arousal, the post-scan ratings of emotional arousal for these pictures were not expected to differ. This was true of the implicit task participants (participants who judged whether pictures occurred indoors or outdoors), who rated moral and non-moral pictures virtually identical [ $F_{1,13} = 0.00, p = 1.00$ ]. However, explicit task participants (participants who rated pictures on moral violation severity) rated moral pictures significantly higher on emotional arousal relative to non-moral pictures [ $F_{1,13} = 14.12, p < .003$ ] (Fig. 3).

Consistent with the higher emotional arousal ratings of moral pictures in participants who performed the explicit task, violation severity ratings of moral pictures were positively correlated with post-scan emotional arousal ratings ( $r(12) = 0.45, p < .0001$ ). Similar correlations were present in the non-moral ( $r(12) = 0.31, p < .0001$ ) and neutral ( $r(12) = 0.23, p < .05$ ) conditions, though the correlation in the moral condition was significantly greater than the non-moral ( $F_{1,12} = 6.23, p < .03$ ) and neutral ( $F_{1,12} = 13.00, p < .005$ ) conditions; the correlations in the non-moral and neutral conditions did not significantly differ ( $F_{1,12} = 1.20, ns$ ). These results suggest that rating pictures on moral violation severity enhanced their perceived emotional salience.

## fMRI results

**Explicit vs. implicit moral picture viewing**—Brain regions in which participants who performed the explicit task showed increased activity during moral relative to non-moral picture viewing are reported by Harenski et al. (2008). Increased activity was present in the ventromedial prefrontal cortex (BA 10/11). Participants who performed the implicit task showed increased activity in bilateral temporo-parietal junction, posterior cingulate (BA 30/31), and, at a reduced statistical threshold, left dorsolateral prefrontal cortex (BA 9;  $p < .002$ , uncorrected). No ventromedial prefrontal activity was present, even at lenient statistical thresholds. At  $p < .05$ , uncorrected, activity was present in an area of medial frontopolar cortex (though this activation, which was 15 mm anterior to the activity observed in participants who performed the explicit task ( $x = 3, y = 60, z = -15$ ), was part of a larger cluster whose peak was located in a more lateral and superior region;  $x = 21, y = 63, z = 0, z$  score = 2.97). Consistent with predictions, between-group comparisons of the moral > non-moral contrast revealed increased activity in ventromedial prefrontal cortex (BA 10/11) in participants who performed the explicit versus the implicit task (Table 1 and Figs. 4A and C). Participants who performed the implicit task showed increased activity in the left



dorsolateral prefrontal cortex relative to participants who performed the explicit task (BA 9; Table 1 and Figs. 4B and D). For a complete list of regions showing increased activity in response to moral vs. non-moral pictures, see Table 1.

The between-group comparison of the moral > neutral contrast also revealed increased ventromedial prefrontal activity (BA 10) in participants who performed the explicit versus the implicit task ( $x = -12, y = 57, z = 0$ ;  $z$  score = 3.01,  $p = .082$ , FWE small-volume corrected). Additional activations were present in left superior frontal cortex ( $x = -24, y = 66, z = 9$  (BA 10);  $z$  score = 2.89) and left premotor cortex ( $x = -21, y = 3, z = 60$  (BA 6);  $z$  score = 3.75). In contrast, the between-group comparison of the non-moral > neutral contrast did not show increased ventromedial prefrontal activity in participants who performed the explicit versus the implicit task; increased activity was present only in left superior frontal cortex ( $x = -30, y = 63, z = 9$  (BA 10);  $z$  score = 3.59). When the explicit moral and implicit moral conditions were directly compared (without reference to the non-moral or neutral conditions), increased ventromedial prefrontal activity was again present ( $x = -3, y = 48, z = -6$  (BA 10/11);  $z$  score = 2.67,  $p = .10$ , FWE small-volume corrected). Multiple activations were also present in frontal, temporal, parietal, and occipital cortex, which is likely due to the fact that the moral rating task was more complex than the indoor/outdoor task.

When violation severity ratings (explicit task) and post-scan emotional arousal ratings (explicit and implicit task) were entered as additional covariates into the same model that was used in the above analyses, the results were unchanged, with the following exceptions: in the explicit moral > non-moral contrast, right temporo-parietal junction activity was present; in the implicit moral > non-moral contrast, posterior cingulate activity was no longer present, and the significance level of the left dorsolateral prefrontal cortex activity was reduced to  $p < .005$ , uncorrected.

**Correlations between brain activity, violation severity ratings, and post-scan emotional arousal ratings**—Violation severity ratings (explicit task) were entered as covariates to investigate whether increased activity in any brain regions during moral picture viewing predicted higher subsequent ratings. A positive correlation was present in the ventromedial prefrontal cortex (BA 11), a region similar to that which showed increased activity during explicit versus implicit moral picture viewing (Fig. 5A). This correlation was not present during non-moral picture viewing. Additional regions positively correlated with violation severity ratings included the posterior cingulate (BA 31) and left dorsolateral prefrontal cortex (BA 9) (see also Harenski et al., 2008).

Post-scan emotional arousal ratings made by participants who performed the explicit task were not significantly correlated with any brain regions during moral picture viewing. In contrast, participants who performed the implicit task showed a positive correlation between medial prefrontal activity (BA 10) and post-scan emotional arousal ratings (Fig. 5B). The activity occurred in a region superior to the ventromedial prefrontal region that showed increased activity during explicit versus implicit moral picture viewing. Positive correlations were also present in the left temporo-parietal junction (BA 39), left dorsolateral prefrontal cortex (BA 9), and right inferior frontal gyrus (BA 47). These correlations were present for

moral pictures only; they did not occur in the non-moral condition. In the non-moral condition, a positive correlation was present in the right anterior insula (BA 13).

**Alternate explanations of increased ventromedial prefrontal activity in explicit vs. implicit moral conditions**—Our hypothesis was that the medial prefrontal cortex would be activated during moral deliberation but not moral intuition. Consistent with this hypothesis, we observed increased ventromedial prefrontal activity during moral picture viewing only in participants who performed the explicit task. However, it is important to consider alternate explanations. One possibility is that the ventromedial prefrontal cortex was actually engaged during moral picture viewing in participants who performed the implicit task, but this activity was suppressed by the cognitive resources required to perform the indoor/outdoor judgment task. If this were the case, we may expect a negative correlation between individual parameter estimates of ventromedial prefrontal activity during moral picture viewing and mean reaction times during the indoor/outdoor judgment. No such correlation was present, however; instead there was a nonsignificant positive correlation ( $r(13) = 0.40, p = .14$ ). Nor was ventromedial prefrontal activity negatively correlated with activity in the dorsolateral prefrontal ( $r(13) = 0.04, p = .89$ ) region that was engaged during moral picture viewing in the implicit condition.

Another possibility concerns potential differences in task difficulty across conditions. The ventromedial prefrontal cortex is part of the brain's 'default network' (Gusnard and Raichle, 2001), a network of regions that often show task-induced deactivation relative to a rest state. Further, the degree of deactivation has been found to correlate with task difficulty (McKiernan et al., 2003). Since our results were due to differences in deactivation (Fig. 4C), it is possible that evaluating unpleasant non-moral pictures for moral content may have constituted a more difficult task than evaluating unpleasant moral pictures for moral content, resulting in deactivation in the non-moral relative to the moral condition.

To test this alternate hypothesis, we used reaction times during the rating period as an index of task difficulty. However, our current task design, in which the participant's reaction times were determined by the rating they gave (increased reaction time for higher ratings due to the continuous progression scale) did not allow us to assess reaction time in a meaningful manner for participants who made violation severity ratings. We thus collected data from a pilot sample of 24 participants (12 female) on the same task outside of the MRI scanner, with the continuous progression scale changed to a 5-point scale with corresponding button presses, and evaluated reaction times associated with violation severity ratings. Reaction times did not significantly differ between moral and non-moral pictures for the entire sample ( $F_{1,22} = 0.10, p = .75$ ), or for the female participants alone ( $F_{1,11} = 0.09, p = .78$ ). Thus, it is unlikely that our results reflect greater task difficulty in the non-moral condition.

## Discussion

The present study contrasted the neural correlates of moral deliberation and moral intuition by comparing two groups of participants who performed either an 'explicit' moral rating task or an 'implicit' indoor/outdoor judgment task. We hypothesized that the medial prefrontal cortex would be more engaged in response to moral pictures during the explicit

relative to the implicit task, and that the temporo-parietal junction would be similarly engaged during both tasks. Both hypotheses were supported, suggesting that the medial prefrontal cortex may contribute more to moral deliberation than moral intuition, whereas the temporo-parietal junction may contribute more to moral intuition than moral deliberation.

Participants performing the explicit but not the implicit task showed increased activity in the ventromedial prefrontal cortex when viewing moral relative to non-moral pictures. This result suggests that the ventromedial prefrontal cortex is more involved in moral deliberation than moral intuition. An important question then is what aspects of moral deliberation engage the ventromedial prefrontal cortex? The explicit moral task included several types of moral deliberation: evaluating pictures for moral content, identifying moral violations, and rating the severity of identified moral violations. Since the evaluation of pictures for moral content occurred in both moral and non-moral conditions, it is unlikely that increased ventromedial prefrontal activity during moral vs. non-moral picture viewing represented the evaluation process. In contrast, the identification and severity rating of moral violations occurred more often during moral relative to non-moral picture viewing; thus the ventromedial prefrontal activity likely contributed to these processes. Indeed, when we entered violation severity ratings from each participant into a regression analysis, we found a positive correlation between ventromedial prefrontal activity during picture viewing and subsequent violation severity ratings in the moral (but not the non-moral) condition.

Participants who performed the implicit task did not show increased ventromedial prefrontal activity during moral vs. non-moral picture viewing, even at the most lenient statistical thresholds ( $p < .05$ , uncorrected), which likely represents the absence of moral deliberation during moral picture viewing. It is possible that viewing moral violations could have engaged moral deliberation, but we do not believe this occurred in our task for two reasons. First, in a prior study involving passive viewing of the same pictures in the present study (Harenski and Hamann, 2006), post-scan interviews indicated that the participants did not engage in moral deliberation, nor were they even aware of the moral/non-moral picture distinction. Second, in the current study participants who performed the implicit task were performing a task that was unrelated to moral deliberation, further decreasing the likelihood that moral deliberation or judgment occurred. As discussed earlier, most prior functional imaging studies of moral processing used an explicit task in which participants were instructed to evaluate the moral appropriateness of stimuli, and reported increased medial prefrontal activity in response to moral stimuli. Whether this activity would still occur if the tasks in these studies were made implicit is an empirical question, though we would speculate that processing complex moral stimuli such as moral dilemmas may be more likely to engage spontaneous moral deliberation than pictures depicting moral violations (and thus medial prefrontal activity might be present).

Participants performing the explicit task showed a positive correlation between ventromedial prefrontal activity during moral picture viewing and violation severity ratings, which were in turn positively correlated with post-scan emotional arousal ratings. We also found that only participants who performed the explicit task rated moral pictures higher on emotional arousal than non-moral pictures. This is interesting given that moral and non-moral pictures

were matched on emotional arousal *a priori*, based on ratings from two separate groups of participants (Harenski and Hamann, 2006; Harenski and Hamann, unpublished pilot data). These results suggest that moral deliberation may have enhanced the perceived emotional arousal of moral pictures, an effect represented by increased ventromedial prefrontal activity. Participants who performed the implicit task showed a positive correlation between medial prefrontal activity and post-scan emotional arousal ratings of moral (but not non-moral) pictures, though this activity was located in a region superior to the ventromedial prefrontal region that was activated in participants performing the explicit task. Thus, it may be that superior medial prefrontal activity represented the inherent emotional salience of moral pictures, whereas ventromedial prefrontal activity represented the interaction of moral deliberation and associated emotional responses.

Participants who performed the implicit task, as well as those who performed the explicit task, showed increased activity in the right temporo-parietal junction during moral relative to non-moral picture viewing. This region has been implicated in a variety of social perception and reasoning processes, including theory of mind (Gallagher and Frith, 2003; Saxe and Kanwisher, 2003), perspective taking, empathy, agency, and self-other distinction (for a review see Decety and Lamm, 2007). Recent work has shown that this region supports intentionality attributions in the context of moral processing (Young et al., 2008). It has also been demonstrated that intentionality cues are processed unconsciously and not accessed during moral deliberation (Cushman et al., 2006). Thus, it is possible that intentionality attributions, represented by right temporo-parietal activity, constitute moral intuitions that occurred during implicit and explicit moral picture viewing in the current study. Although it would be reasonable to suggest that these processes also contributed to moral deliberation, our results do not support this suggestion, since if this were the case we should have observed increased temporo-parietal activity in the explicit versus implicit moral task, which we did not.

Implicit, but not explicit, task participants showed increased dorsolateral prefrontal activity during moral vs. non-moral picture viewing. This region has not been consistently implicated in moral processing in prior studies; however, dorsolateral prefrontal activity has been reported in specific moral processing contexts, including utilitarian vs. non-utilitarian moral judgments and difficult vs. easy moral judgments (Greene et al., 2001, 2004), and has also been shown to correlate with individual moral competence (Prehn et al., 2008). Unlike the current study, these studies highlighted a role of the dorsolateral prefrontal cortex in moral deliberation. In the current study we observed increased dorsolateral prefrontal activity only in the implicit moral condition, where no moral deliberation occurred. Additionally, the dorsolateral prefrontal activity in the studies mentioned above occurred in different regions than the one activated in the current study (Brodmann areas 10, 45, and 46 versus Brodmann area 9). Thus, the dorsolateral prefrontal activity in the current study may represent different underlying processes than prior studies. One possibility is that the presence of moral content was salient enough to compete with the attentional resources required to make indoor/outdoor discriminations, and the increased dorsolateral prefrontal activity represented conflict associated with the concurrent processing of task-relevant and salient-but-task-irrelevant information. Even if moral deliberation was not occurring,

attention to moral cues (e.g., a pointed gun) may have created some form of interference. Functional imaging research utilizing the Stroop task has shown that the region of dorsolateral prefrontal cortex identified in the current study is involved in cognitive control, i.e., maintaining task performance in the presence of irrelevant information (MacDonald et al., 2000). This region is also known to be involved in working memory, particularly the monitoring and manipulation of information held in working memory (D'Esposito et al., 2000).

In light of the present results, which show increased ventromedial prefrontal activity in participants who performed the explicit moral task but not the implicit moral task, it is important to note the results of a recent study that did report increased ventromedial prefrontal activity during implicit moral processing (Moll et al., 2007). This study found that reading statements evoking prosocial moral emotions (e.g., guilt and compassion) increased activity in a ventromedial prefrontal region similar to the one that we observed during explicit relative to implicit moral processing (see also Kedia et al., 2008; Zahn et al., 2009). Since our ventromedial prefrontal activity was found to be associated with increased emotional responses during moral deliberation, perhaps these represent specifically prosocial emotional responses. When compared with our finding of an association between emotional responses to moral stimuli and activity in a more superior medial prefrontal region in participants who performed the implicit task, an intriguing possibility is that explicit and implicit processing of the same moral stimuli evoke different types of moral emotions. The findings of Moll et al. (2007) could also indicate that ventromedial prefrontal activity is engaged during moral intuition (in the absence of moral deliberation) under certain conditions but not others (e.g., when specific moral emotions are elicited). However, this cannot be concluded unless it is clear that participants in Moll et al. did not engage in spontaneous moral deliberation while reading moral statements. As discussed earlier, passive viewing pictures depicting moral violations may be less likely to engage moral deliberation (especially when participants are performing a task unrelated to moral deliberation) relative to reading statements or complex scenarios describing moral violations. This point underscores the importance of post-task debriefing where participants are asked whether they engaged in moral deliberation during implicit processing of moral stimuli.

Although thus far we have described moral intuition as preceding moral deliberation, it may also be possible for moral deliberation to engender moral intuitions. If this did occur in the current study it is unlikely that the intuitions are those represented by the right temporal-parietal junction, since activity in this region occurred during the explicit and implicit tasks, the latter which did not involve moral deliberation. The increased emotional responses to moral stimuli that occurred in the explicit task may represent affective intuitions that are generated by moral deliberation (and represented by increased ventromedial prefrontal activity).

To further explore the different types of processing that occur during moral deliberation and intuition, it will be helpful for future studies to determine whether explicit vs. implicit tasks promote processing of different features of moral stimuli. Pictorial moral stimuli are complex and include multiple cues to moral violations such as facial expressions, body

gestures, weapons or other objects. Certain features of moral stimuli might draw attention for the purpose of evaluating moral salience in an explicit moral task, whereas different features might draw attention during an implicit task. Indoor/outdoor judgments might shift focus away from certain aspects of moral stimuli to a greater extent than others. This could be assessed by eye tracking, or memory testing for different aspects of moral pictures.

A limitation of the present study is that we did not obtain post-scan violation severity ratings from participants who performed the implicit task. Violation severity ratings were included only in the model of participants who performed the explicit task. This could result in a better fit of their model relative to the participants who performed the implicit task, since these ratings reflect the personal values of each participant.

To our knowledge, this study is the first to directly compare the neural correlates of moral deliberation and moral intuition. We observed overlapping and distinct patterns of activation during each type of moral processing, in the temporo-parietal junction and medial and lateral prefrontal cortices, respectively. The identification of neural networks engaged in moral deliberation and intuition may be useful in studies investigating groups of individuals that show deficits moral processing. For example, Koenigs et al. (2007) found that patients with medial prefrontal damage made more utilitarian moral judgments relative to patients without damage to this region, indicating a deficit that affected moral deliberation (and possibly moral intuition). Another group of individuals that is receiving increasing interest in this domain are psychopaths, who exhibit personality and behavioral characteristics such as callousness, manipulateness, and criminal behavior, and are generally considered to be deficient in moral sensitivity. de Oliveira-Souza et al. (2008) found that psychopaths, relative to non-psychopaths, showed widespread gray matter reductions in several brain regions comprising the ‘moral neural network,’ including the medial prefrontal cortex, temporo-parietal junction and anterior temporal cortex. Whether this dysfunction in the neural circuitry that contributes to moral processing is related to deficits in moral intuition and/or deliberation is a question for future research examining both types of processes in these and other groups of individuals with known neural dysfunction or deficits in moral processing.

## Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

## References

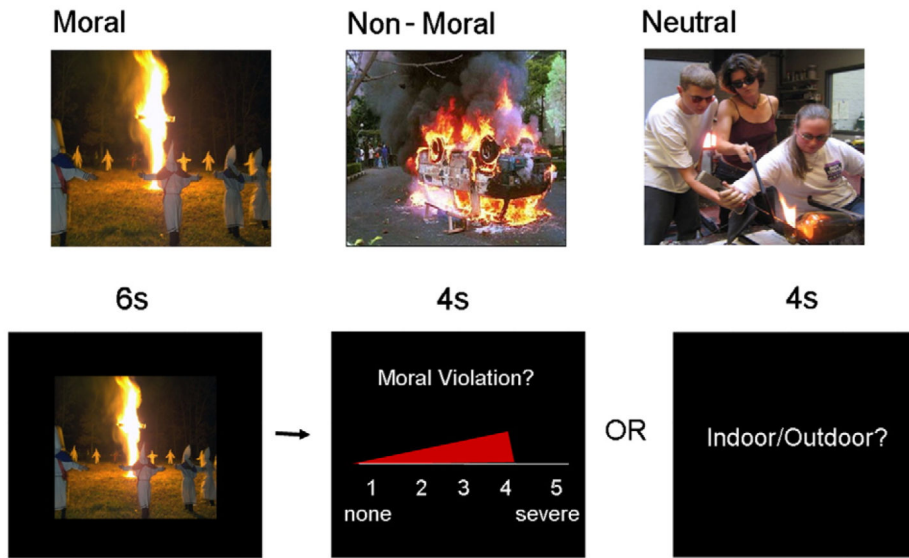
- Amodio DM, Frith CD. Meeting of minds: the medial frontal cortex and social cognition. *Nature Reviews Neuroscience*. 2006; 7:268–277.
- Bechara A, Damasio H, Tranel D, Damasio AR. Deciding advantageously before knowing the advantageous strategy. *Science*. 1997; 275:1293–1295. [PubMed: 9036851]
- Bechara A, Damasio H, Damasio AR. Different contributions of the human amygdala and ventromedial prefrontal cortex to decision-making. *J. Neurosci*. 1999; 19:5473–5481. [PubMed: 10377356]
- Bechara A, Damasio H, Damasio AR. Emotion, decision making and the orbitofrontal cortex. *Cereb. Cortex*. 2000; 10:295–307. [PubMed: 10731224]

- Berthoz S, Grezes J, Armony JL, Passingham RE, Dolan RJ. Affective response to one's own moral violations. *Neuroimage*. 2006; 31:945–950. [PubMed: 16490367]
- Ciaramidaro A, Adenzato M, Enrici I, Erk S, Pia L, Bara BG, Walter H. The intentional network: how the brain reads varieties of intentions. *Neuropsychologia*. 2007; 45:3105–3113. [PubMed: 17669444]
- Cunningham WA, Johnson MK, Gatenby JC, Gore JC, Banaji MR. Neural components of social evaluation. *J. Pers. Soc. Psychol.* 2003; 85:639–649. [PubMed: 14561118]
- Cushman F, Young L, Hauser M. The role of conscious reasoning and intuition in moral judgment: testing three principles of harm. *Psychol. Sci.* 2006; 17:1082–1089. [PubMed: 17201791]
- Damasio AR, Grabowski TJ, Bechara A, Damasio H, Ponto LLB, Parvizi J, Hichwa RD. Subcortical and cortical brain activity during the feeling of self-generated emotions. *Nat. Neurosci.* 2000; 3:1049–1056. [PubMed: 11017179]
- Decety J, Lamm C. The role of the right temporoparietal junction in social interaction: how low-level computational processes contribute to meta-cognition. *Neurosci.* 2007; 13:580–593.
- D'Esposito M, Postle B, Rypma B. Prefrontal cortical contributions to working memory: evidence from event-related fMRI studies. *Exp. Brain Res.* 2000; 133:3–11. [PubMed: 10933205]
- de Oliveira-Souza R, Hare RD, Bramati IE, Garrido GJ, Ignacio FA, Tovar-Moll F, Moll J. Psychopathy as a disorder of the moral brain: fronto-temporo-limbic grey matter reductions demonstrated by voxel-based morphometry. *NeuroImage*. 2008; 40:1203–1213.
- Fellows LK, Farah MJ. Different underlying impairments in decision-making following ventromedial and dorsolateral prefrontal damage in humans. *Cereb. Cortex*. 2005; 15:58–63. [PubMed: 15217900]
- Fellows LK, Farah MJ. The role of ventromedial prefrontal cortex in decision making: judgment under uncertainty or judgment per se? *Cereb. Cortex*. 2007; 17:2669–2674. [PubMed: 17259643]
- Fink GR, Markowitsch HJ, Reinkemeier M, Bruckbauer T, Kessler J, Heiss WD. Cerebral representation of one's own past: neural networks involved in autobiographical memory. *J. Neurosci.* 1996; 16:4275–4282. [PubMed: 8753888]
- Friston KJ, Jezzard P, Turner R. Analysis of functional MRI time-series. *Hum. Brain Mapp.* 1994; 1:153–171.
- Gallagher HL, Frith CD. Functional imaging of 'theory of mind'. *Trends Cogn. Sci.* 2003; 7(2):77–83. [PubMed: 12584026]
- Greene J, Haidt J. How (and where) does moral judgment work? *Trends Cogn. Sci.* 2002; 6:517–523. [PubMed: 12475712]
- Greene JD, Sommerville RB, Nystrom LE, Darley JM, Cohen JD. An fMRI investigation of emotional engagement in moral judgment. *Science*. 2001; 293:2105–2108. [PubMed: 11557895]
- Greene JD, Nystrom LE, Engell AD, Darley JM, Cohen JD. The neural bases of cognitive conflict and control in moral judgment. *Neuron*. 2004; 44:389–400. [PubMed: 15473975]
- Gusnard DA, Raichle ME. Searching for a baseline: functional imaging and the resting human brain. *Nature Reviews Neuroscience*. 2001; 2:685–694.
- Haidt J, Koller S, Dias M. Affect, culture, and morality, or is it wrong to eat your dog? *J. Pers. Soc. Psychol.* 1993; 65:613–628. [PubMed: 8229648]
- Haidt J. The emotional dog and its rational tail: a social intuitionist approach to moral judgment. *Psychol. Rev.* 2001; 108(4):814–834. [PubMed: 11699120]
- Harenski CL, Hamann S. Neural correlates of regulating negative emotions related to moral violations. *NeuroImage*. 2006; 30:313–324. [PubMed: 16249098]
- Harenski CL, Antonenko O, Shane M, Kiehl K. Gender differences in neural mechanisms underlying moral sensitivity. *Soc. Cogn. Affect. Neurosci.* 2008; 3(4):313–321. [PubMed: 19015084]
- Hauser MD. The liver and the moral organ. *Soc. Cogn. Affect. Neurosci.* 2006; 1:214–220. [PubMed: 18985108]
- Heekeren HR, Wartenburger I, Schmidt H, Schwintowski HP, Villringer A. An fMRI study of simple ethical decision-making. *NeuroReport*. 2003; 14:1215–1219. [PubMed: 12824762]

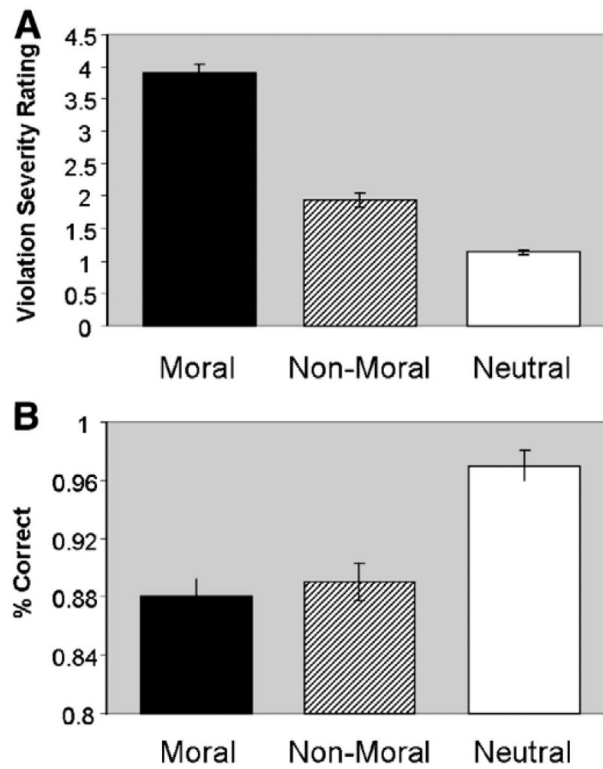
- Heekeren HR, Wartenburger I, Schmidt H, Prehn K, Schwintowski HP, Villringer A. Influence of bodily harm on neural correlates of semantic and moral decision making. *NeuroImage*. 2005; 24:887–897. [PubMed: 15652323]
- Holmes AP, Josephs O, Buchel C, Friston KJ. Statistical modelling of low frequency confounds in fMRI. *NeuroImage*. 1997; 5:S480.
- Johnson MK, Raye CL, Mitchell KM, Touryan SR, Greene EJ, Nolen-Hoeksema S. Dissociating medial frontal and posterior cingulate activity during self reflection. *Soc. Cogn. Affect. Neurosci*. 2006; 1:56–64. [PubMed: 18574518]
- Kedia G, Berthoz S, Wessa M, Hilton D, Martinot J. An agent harms a victim: a functional magnetic resonance imaging study on specific moral emotions. *J. Cogn. Neurosci*. 2008; 20:1788–1798. [PubMed: 18211233]
- Koenigs M, Young L, Adolphs R, Tranel D, Cushman F, Hauser M, Damasio A. Damage to the prefrontal cortex increases utilitarian moral judgments. *Nature*. 2007; 446:908–911. [PubMed: 17377536]
- Lang, P.J.; Bradley, M.M.; Cuthbert, B.N. International Affective Picture System (IAPS). National Institute of Mental Health Center for the Study of Emotion and Attention; Bethesda, MD: 1995.
- MacDonald AW III, Cohen JD, Stenger VA, Carter CS. Dissociating the role of the dorsolateral prefrontal and anterior cingulate cortex in cognitive control. *Science*. 2000; 288:1835–1838. [PubMed: 10846167]
- McKiernan KA, Kaufman JN, Kucera-Thompson J, Binder JR. A parametric manipulation of factors affecting task-induced deactivation in functional neuroimaging. *Journal of Cognitive Neuroscience*. 2003; 15:394–408. [PubMed: 12729491]
- Maddock RJ. The retrosplenial cortex and emotion: new insights from functional neuroimaging of the human brain. *Trends Neurosci*. 1999; 22:310–316. [PubMed: 10370255]
- Moll J, de Oliveira-Souza R, Bramati IE, Grafman J. Functional networks in emotional moral and non-moral judgments. *NeuroImage*. 2002a; 16:696–703. [PubMed: 12169253]
- Moll J, de Oliveira-Souza R, Eslinger PJ, Bramati IE, Mourao-Miranda J, Andreiuolo PA, Pessoa L. The neural correlates of moral sensitivity: a functional magnetic resonance imaging investigation of basic and moral emotions. *J. Neurosci*. 2002b; 22(7):2730–2736. [PubMed: 11923438]
- Moll J, Zahn R, de Oliveira-Souza R, Krueger F, Grafman J. The neural basis of human moral cognition. *Nat. Rev., Neurosci*. 2005; 6:799–809. [PubMed: 16276356]
- Moll J, de Oliveira-Souza R, Garrido GJ, Bramati IE, Caparelli-Daquer EMA, Paiva MLMF, Zahn R, Grafman J. The self as a moral agent: linking the neural bases of social agency and moral sensitivity. *Soc. Neurosci*. 2007; 2:336–352. [PubMed: 18633822]
- Paulus MP, Frank LR. Ventromedial prefrontal cortex activation is critical for preference judgments. *NeuroReport*. 2003; 14:1311–1315. [PubMed: 12876463]
- Pizarro DA, Bloom P. The intelligence of the moral intuitions: a reply to Haidt. *Psychol. Rev*. 2003; 110:193–196. [PubMed: 12529062]
- Prehn K, Wartenburger I, Meriau K, Scheibe C, Goodenough OR, Villringer A, van der Meer E, Heekeren H. Individual differences in moral judgment competence influence neural correlates of socio-normative judgments. *Soc. Cogn. Affect. Behav. Neurosci*. 2008; 3:33–46.
- Robertson D, Snarey J, Ousley O, Harenski K, Bowman FD, Gilkey R, Kilts CK. The neural processing of moral sensitivity to issues of justice and care. *Neuropsychologia*. 2007; 45:755–766. [PubMed: 17174987]
- Ruby P, Decety J. What you believe versus what you think they believe: a neuroimaging study of conceptual perspective-taking. *Eur. J. Neurosci*. 2003; 17:2475–2480. [PubMed: 12814380]
- Sanfey AG, Hastie R, Colvin MK, Grafman J. Phineas gauged: decision-making and the human prefrontal cortex. *Neuropsychologia*. 2003; 41:1218–1229. [PubMed: 12753961]
- Saxe R, Kanwisher N. People thinking about people: the role of the temporoparietal junction in ‘theory of mind’. *NeuroImage*. 2003; 19:1835–1842. [PubMed: 12948738]
- Schaich Borg J, Hynes C, Van Horn J, Grafton S, Sinnott-Armstrong W. Consequences, action, and intention as factors in moral judgments: an fMRI investigation. *J. Cogn. Neurosci*. 2006; 18:803–817. [PubMed: 16768379]



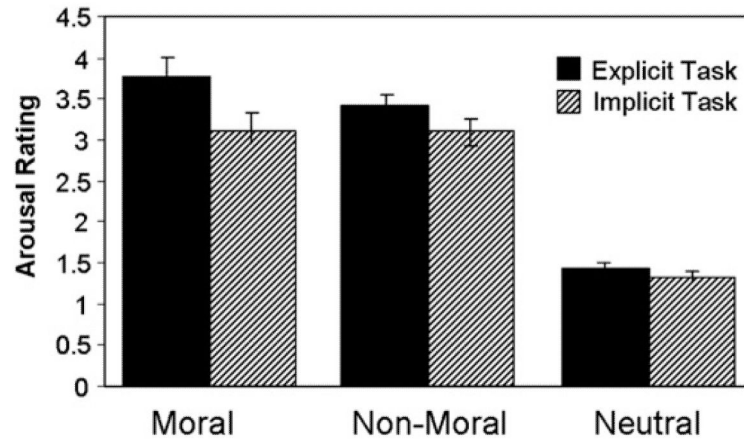
- Schaich Borg J, Lieberman D, Kiehl KA. Infection, incest, and iniquity: investigating the neural correlates of disgust and morality. *J. Cogn. Neurosci.* 2008; 20:1529–1546. [PubMed: 18345982]
- Schweder, RA.; Mahapatra, M.; Miller, J. Culture and moral development. In: Kagan, J.; Lamb, S., editors. *The emergence of morality in young children.* University of Chicago Press; Chicago: 1987. p. 1-83.
- Vogt, BA.; Laureys, S. Posterior cingulate, precuneal and retrosplenial cortices: cytology and components of the neural network correlates of consciousness. In: Laureys, S., editor. *Progress in Brain Research.* Vol. 150. Elsevier; 2005. p. 205-217.
- Young L, Saxe R. The neural basis of belief encoding and integration in moral judgment. *Neuroimage.* 2008; 40:1912–1920. [PubMed: 18342544]
- Young L, Cushman F, Hauser M, Saxe R. The neural basis of the interaction between theory of mind and moral judgment. *Proc. Natl. Acad. Sci.* 2007; 104:8235–8240. [PubMed: 17485679]
- Zahn R, Moll J, Paiva M, Garrido G, Krueger F, Huey ED, Grafman J. The neural basis of human social values: evidence from functional MRI. *Cereb. Cortex.* 2009; 19:276–283. [PubMed: 18502730]



**Fig. 1.** (A) Example moral, non-moral, and neutral pictures. (B) Example 'moral trial' for the explicit and implicit tasks.

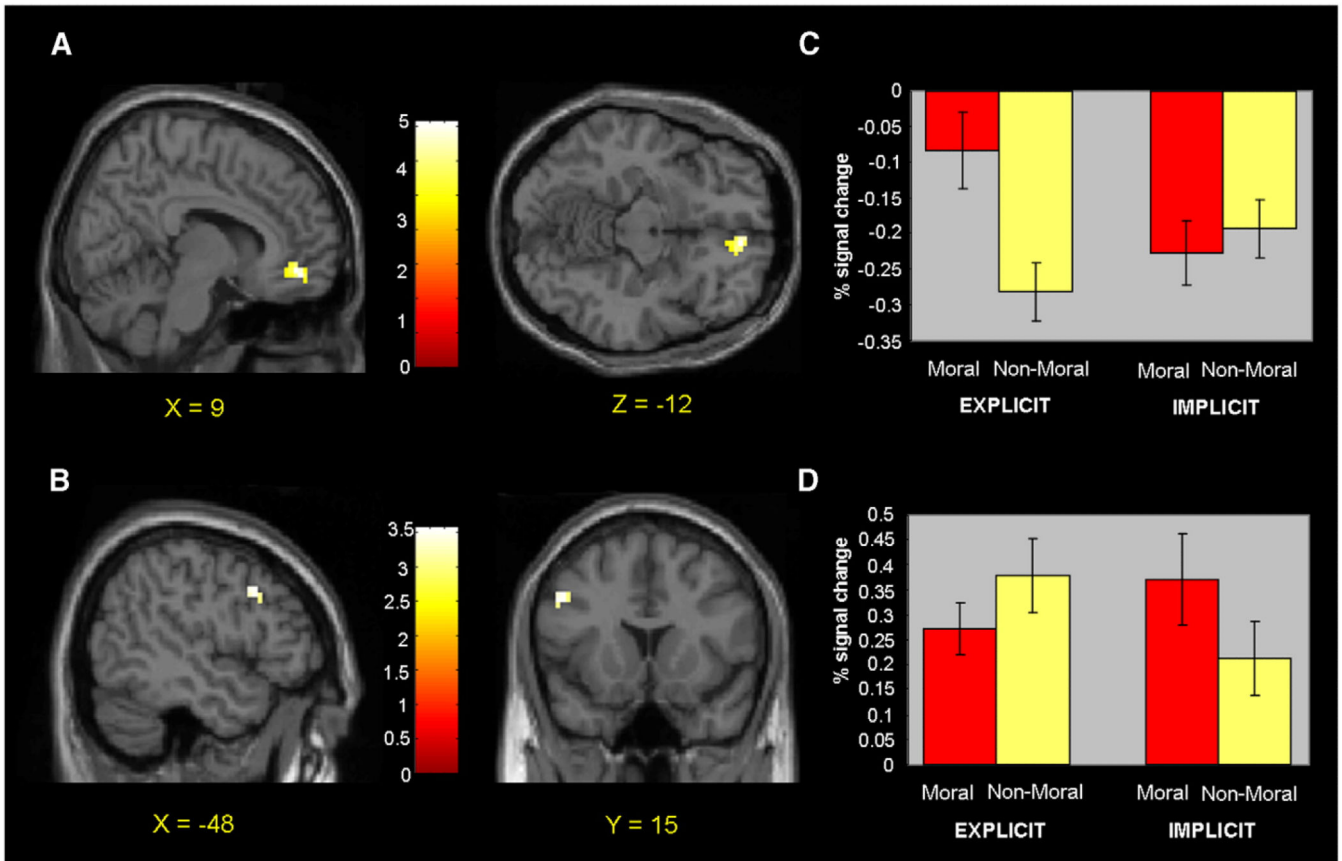


**Fig. 2.** (A) Moral violation severity ratings by condition, indicating higher violation severity ratings in response to moral relative to non-moral and neutral pictures (Harenski et al., 2008). (B) Accuracy of indoor/outdoor ratings by condition, indicating higher accuracy for neutral relative to moral and non-moral pictures, but no significant difference for moral vs. non-moral pictures.

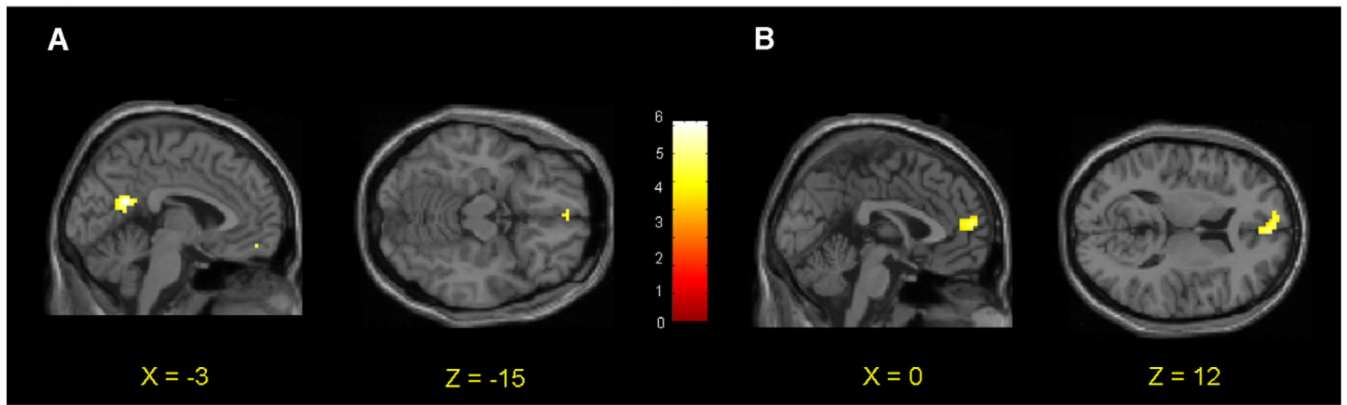


**Fig. 3.**

Post-scan arousal ratings across conditions for explicit and implicit task participants. Both explicit and implicit task participants rated moral and non-moral pictures significantly higher on emotional arousal relative to neutral pictures. Participants who performed the explicit task rated moral pictures significantly higher on arousal relative to non-moral pictures, whereas no such effect was present in participants who performed the implicit task.



**Fig. 4.** (A) Increased ventromedial prefrontal activity during moral picture viewing in explicit vs. implicit task participants. (B) Increased dorsolateral prefrontal activity during moral picture viewing in implicit vs. explicit task participants. (C) Percent signal change values for ventromedial prefrontal activity across moral and non-moral condition in explicit and implicit task participants. (D) Percent signal change values for dorsolateral prefrontal activity across moral and non-moral condition in explicit and implicit task participants.



**Fig. 5.**

(A) Positive correlation between ventromedial prefrontal activity (BA 10/11) and violation severity ratings of moral pictures in participants who performed the explicit task. (B) Positive correlation between medial prefrontal activity (BA 10) and post-scan emotional arousal ratings of moral pictures in participants who performed the implicit task.

**Table 1**

Brain activity during moral picture viewing in explicit vs. implicit task participants.

Moral > Non-Moral	BA	Explicit task		Implicit task		Explicit > Implicit		Implicit > Explicit	
		<i>z</i> ( <i>k</i> )	<i>p</i> value	<i>z</i> ( <i>k</i> )	<i>p</i> value	<i>z</i> ( <i>k</i> )	<i>p</i> value	<i>z</i> ( <i>k</i> )	<i>p</i> value
		( <i>x,y,z</i> )		( <i>x,y,z</i> )		( <i>x,y,z</i> )		( <i>x,y,z</i> )	
<i>Regions of interest</i>									
R. medial frontal gyrus	10/11	4.70 (158) (12,45,-9)	.001	–	–	4.28 (137) (9,48,-12)	.001	–	–
R. medial frontal gyrus	10	–	–	–	–	–	–	–	–
L. temporo-parietal junction	39	–	–	3.83 (56) (-45,-72,24)	.025	–	–	–	–
L. temporo-parietal junction	39	–	–	3.67 (38) (-45,-72,24)	.046	–	–	–	–
R. temporo-parietal junction	39	3.23 (22) <sup>a</sup> (48,-72,39)	.030	3.85 (72) (48,-78,27)	.013	–	–	–	–
R. posterior cingulate	30	–	–	3.94 (57) (9,-54,9)	.011	–	–	–	–
<i>Other regions</i>									
L. parahippocampal gyrus	36	–	–	3.83 (16) (-27,-45,-15)	.001	–	–	–	–
L. cerebellum		–	–	3.67 (5) (-15,-36,-12)	.001	–	–	–	–
L. middle temporal gyrus	21	–	–	4.16 (9) (-63,-39,-9)	.001	–	–	–	–
L. middle Occipital occipital gyrus	19	–	–	3.41 (75) (-54,-75,6)	.001	–	–	–	–
L. middle frontal gyrus	9	–	–	3.06 (40) (-48,12,39)	.002	–	–	2.98 (5) (-48,15,33)	.005

BA = Brodmann area,  $Z=z$  score,  $k$  = spatial extent of activation. Activations are reported in MNI coordinate space.  $p$  values for regions of interest represent FWE small-volume-corrected values;  $p$  values for other regions are uncorrected. When all explicit and implicit participants were analyzed together, two additional activations (not present in either group when analyzed separately) were present in left posterior cingulate (BA 31,  $x = -6$ ,  $y = -54$ ,  $z = 24$ ;  $z$  score = 4.04) and right anteromedial prefrontal cortex (BA 10,  $x = 3$ ,  $y = 60$ ,  $z = 6$ ;  $z$  score = 3.70).

<sup>a</sup> Activity was present when violation severity ratings were included in the model. When ratings were not included in the model, activity was present at  $p < .002$ , uncorrected.