

# The Neural Correlates of Moral Sensitivity: A Functional Magnetic Resonance Imaging Investigation of Basic and Moral Emotions

Jorge Moll,<sup>1</sup> Ricardo de Oliveira-Souza,<sup>1</sup> Paul J. Eslinger,<sup>2</sup> Ivanei E. Bramati,<sup>1</sup> Janaína Mourão-Miranda,<sup>1,3</sup> Pedro Angelo Andreiuolo,<sup>1</sup> and Luiz Pessoa<sup>4</sup>

<sup>1</sup>Neuroimaging and Behavioral Neurology Group, Hospitais D'Or and LABS, RJ, 22281-081, Brazil, <sup>2</sup>Departments of Medicine (Division of Neurology), Pediatrics, and Behavioral Science, Pennsylvania State University College of Medicine, The Milton S. Hershey Medical Center, Hershey, Pennsylvania 17033, <sup>3</sup>Carlos Chagas Filho Biophysics Institute, Federal University of Rio de Janeiro, Rio de Janeiro, RJ, 21941, Brazil, and <sup>4</sup>Laboratory of Brain and Cognition, National Institute of Mental Health, Bethesda, Maryland 20892

Humans are endowed with a natural sense of fairness that permeates social perceptions and interactions. This moral stance is so ubiquitous that we may not notice it as a fundamental component of daily decision making and in the workings of many legal, political, and social systems. Emotion plays a pivotal role in moral experience by assigning human values to events, objects, and actions. Although the brain correlates of basic emotions have been explored, the neural organization of “moral emotions” in the human brain remains poorly understood. Using functional magnetic resonance imaging and a passive visual task, we show that both basic and moral emotions activate the amygdala, thalamus, and upper midbrain. The

orbital and medial prefrontal cortex and the superior temporal sulcus are also recruited by viewing scenes evocative of moral emotions. Our results indicate that the orbital and medial sectors of the prefrontal cortex and the superior temporal sulcus region, which are critical regions for social behavior and perception, play a central role in moral appraisals. We suggest that the automatic tagging of ordinary social events with moral values may be an important mechanism for implicit social behaviors in humans.

*Key words:* moral judgment; fMRI; emotion; orbitofrontal; sociopathy; frontal lobes

Recent theoretical developments in moral psychology, which had been dominated by rationalistic theories for centuries, have emphasized the role of emotion in models of moral development and behavior (Gilligan, 1993; Haidt, 2001). “Moral emotions” have been the focus of several recent experimental psychology studies; they differ from basic emotions in that they are intrinsically linked to the interests or welfare either of society as a whole or of persons other than the agent (Damasio, 1994; de Waal, 1996; Rozin et al., 1999; Haidt, 2002). Moral emotions are readily evoked by the perception of moral violations; it has been suggested that, in contrast to laborious deductive reasoning, they enable rapid, automatic, and unconscious cognitive appraisals of interpersonal events (Haidt, 2001).

Impaired socio-moral emotion and behavior has been consistently observed in patients with frontal-lobe dysfunction, and may dissociate from social cognition and moral knowledge (Eslinger and Damasio, 1985; Price et al., 1990; Eslinger et al., 1992; Anderson et al., 1999). Factors such as the age of the patient at the time of the injury and the location of frontal-lobe damage appear to influence whether social and moral cognition are also impaired. Analogous dissociations between knowing how to behave and actually behaving in socially desirable ways are observed in developmental sociopaths as well (Blair, 1995). In addition, these individuals show a disproportionate impairment in the

ability to experience certain moral emotions (Blair et al., 1995). Moreover, emotion-induced reflex epileptic seizures triggered specifically by feelings related to unresolved moral issues have been reported in association with right frontotemporal damage (Cohen et al., 1999). Together, these findings suggest the existence of brain networks specialized for the generation of moral emotions.

Here we investigate the neural correlates of moral emotion in normal individuals using functional magnetic resonance imaging (fMRI). Because damage to the orbitofrontal cortex (OFC) can result in a lack of empathy and antisocial behaviors in the face of preserved social cognition and basic emotions (Eslinger and Damasio, 1985; Saver and Damasio, 1991; Eslinger et al., 1992), we hypothesized that this brain region would be more activated by the visual perception of stimuli evocative of moral emotions compared with emotional stimuli without moral content and compared with non-emotional stimuli. In addition, we predicted that basic and moral emotions would evoke overlapping activations in brain regions involved in emotional processing, such as the amygdala, insula, and subcortical nuclei (Damasio, 1994; Lane et al., 1999; Damasio et al., 2000).

## MATERIALS AND METHODS

**Subjects.** Seven right-handed adults [five men and two women; age,  $27 \pm 3$  years; education,  $14 \pm 2$  years (results expressed as means  $\pm$  SDs)] with no history of neurological or psychiatric disorders were studied. Subjects were not paid for their participation. The study was conducted at Hospital Barra D'Or and approved by the institutional review boards and ethics committees of the hospital.

**Stimuli and task.** Subjects were scanned while viewing pictures of emotionally charged scenes with and without moral content as well as emotionally neutral pictures. Because our goal was to investigate spon-

Received Nov. 16, 2001; revised Jan. 7, 2002; accepted Jan. 8, 2002.

We thank J. Grafman for valuable discussions and J. Weisberg for comments on drafts of this manuscript.

Correspondence should be addressed to Jorge Moll, Neuroimaging and Behavioral Neurology Group, Hospitais D'Or and LABS, R. Pinheiro Guimarães 22, 4th floor, Rio de Janeiro, RJ, 22281-081, Brazil. E-mail: jmoll@neuroimage.com.

Copyright © 2002 Society for Neuroscience 0270-6474/02/222730-07\$15.00/0

**Table 1. Ratings for emotional valence, arousal, and moral content**

Dimensions	Conditions				
	Interesting	Pleasant	Unpleasant	Neutral	Moral
Valence <sup>a</sup>	6.53 ± 2.49	7.97 ± 1.51	2.40 ± 1.91	5.63 ± 1.42	4.53 ± 2.14
Arousal <sup>b</sup>	6.06 ± 1.91	6.29 ± 2.30	7.01 ± 2.07	4.90 ± 2.38	6.35 ± 2.07
Moral <sup>c</sup>	3.99 ± 2.46	4.50 ± 2.79	4.24 ± 2.64	3.73 ± 2.47	7.13 ± 2.15

<sup>a</sup>  $F_{(5,858)} = 181.03$ ;  $p < 0.0001$ .

<sup>b</sup>  $F_{(5,858)} = 30.48$ ;  $p < 0.0001$ .

<sup>c</sup>  $F_{(5,858)} = 42.61$ ;  $p < 0.001$ .

taneous brain responses triggered by the perception of visual stimuli, no responses were required from the subjects during scanning; they were simply instructed to view the pictures attentively. There were six experimental conditions: (1) moral pictures portraying emotionally charged, unpleasant social scenes, representing moral violations (e.g., physical assaults, poor children abandoned in the streets, war scenes) (Haidt, 2002); (2) unpleasant pictures of aversive scenes not conveying moral connotations (e.g., body lesions, dangerous animals, body products) (Rozin et al., 1999); (3) pleasant pictures, including scenes of people and landscapes; (4) “interesting” pictures, which were visually arousing but less emotional (e.g., surreal images, radical sports); (5) neutral pictures, including people and landscapes; and (6) scrambled images. Stimuli pertaining to these categories were selected based on previous piloting behavioral observations in a different group of normal subjects. After fMRI scanning, subjects rated each picture for moral content, emotional valence, and level of arousal on visual analog scales. Images were presented through magnetically shielded liquid crystal display goggles (Resonance Technologies, Northridge, CA) controlled by a computer. A blocked design consisting of three pictures per block, each displayed for 5 sec, was used. Eight blocks for each category were presented in a fixed pseudorandomized order. A white cross-hair fixation on a black background was displayed for 15 sec between each block to allow for a complete return of the hemodynamic response to baseline levels. Most pictures were selected from the International Affective Picture System (Lang et al., 1995). All pictures were presented again immediately after the imaging session, and subjects rated them according to predefined criteria. The rating dimensions, which were explained after fMRI data acquisition, included emotional valence (from “extremely unpleasant” to “extremely pleasant”), level of arousal (from “calm” to “excited”), and moral content (from “absent” to “extreme”). Subjects were instructed to rate the pictures on these dimensions according to their appraisal of the pictures when first viewing them while in the scanner (1 corresponded to extremely unpleasant/calm/absent and 10 corresponded to extremely pleasant/excited/extreme). Explicit written instructions regarding the rating of moral pictures were provided: “This item refers to the presence and degree of moral content in the scene. This includes actions which you consider to be commendable or regrettable, fair or unfair, right or wrong, good or evil, or situations that evoke a sense of friendship, betrayal, pity or care for others, humiliation, gratitude, or indignation.”

**Behavioral data analysis.** Mean ratings for moral content, emotional valence, and arousal in each experimental condition were assessed by one-way analysis of covariance. Pairwise comparisons of means between conditions were evaluated *post hoc* with Scheffé’s test. A threshold of significance of  $p < 0.05$  was adopted for all comparisons.

**fMRI procedures.** Anatomic data consisted of volumetric  $T_1$ -weighted gradient-echo images [repetition time (TR)/echo time (TE) of 14/4.6 msec; matrix of  $256 \times 256$ ; field of view (FOV) of 256 mm; thickness/gap of 1.25/0 mm;  $n = 128$  slices]. Functional data were acquired with blood oxygen level dependent contrast-echoplanar imaging (TR/TE of 4980/66 msec; matrix of  $128 \times 128$ ; FOV of 256 mm; thickness/gap of 5/0.25 mm;  $n = 16$  slices). All imaging data were obtained with a 1.5 tesla MRI scanner (Vision; Siemens, Erlangen, Germany) equipped with a high-performance gradient overdrive system. Warping of the echoplanar data was minimized by the use of a fast gradient-switching system and by carefully performing a three-dimensional localized shimming procedure before image acquisition in each subject. The use of a higher-resolution matrix ( $128 \times 128$  instead of  $64 \times 64$ ), maintaining a small FOV, also helps to reduce image distortion at the base of the brain in echoplanar imaging. As reported previously (Moll et al., 2001), these imaging parameters allowed adequate signal-to-noise levels in critical brain regions, such as the anterior and inferior prefrontal cortex and medial temporal

lobe (including the amygdala), which was confirmed by carefully inspecting the raw echoplanar data sets. Severe signal dropouts (>50%) were observed in only a small part of the midlateral basal temporal lobe (adjacent to the petrous temporal bone and mastoid cells) and in the posterior portion of the medial basal/subgenual frontal cortex. Therefore, the role of these brain regions in the brain processes investigated here may have been underestimated, a ubiquitous issue in current fMRI investigations. Functional data sets were motion-corrected for three dimensions, and slice time correction, temporal smoothing, linear trend removal, and spatial smoothing (isotropic Gaussian kernel, 6 mm full width at half maximum) were performed. Data sets were coregistered and Talairach-transformed (Talairach and Tournoux, 1988). Activation maps were analyzed using statistical parametric methods (Friston et al., 1995) contained in the BrainVoyager version 4.4 software (Brain Innovation, Maastricht, The Netherlands). Data acquisition was synchronized with the stimulus presentation. Regressors representing the experimental conditions of interest were modeled with a hemodynamic response filter and entered into a multiple-regression analysis using a fixed-effects model. Contrasts between conditions of interest were assessed with  $t$  statistics. A minimum three-dimensional cluster extent threshold of  $50\text{--}100\text{ mm}^3$  was used to protect against type I errors (Forman et al., 1995). Statistical parametric maps were overlaid on two- and three-dimensional renderings of a representative individual’s brain.

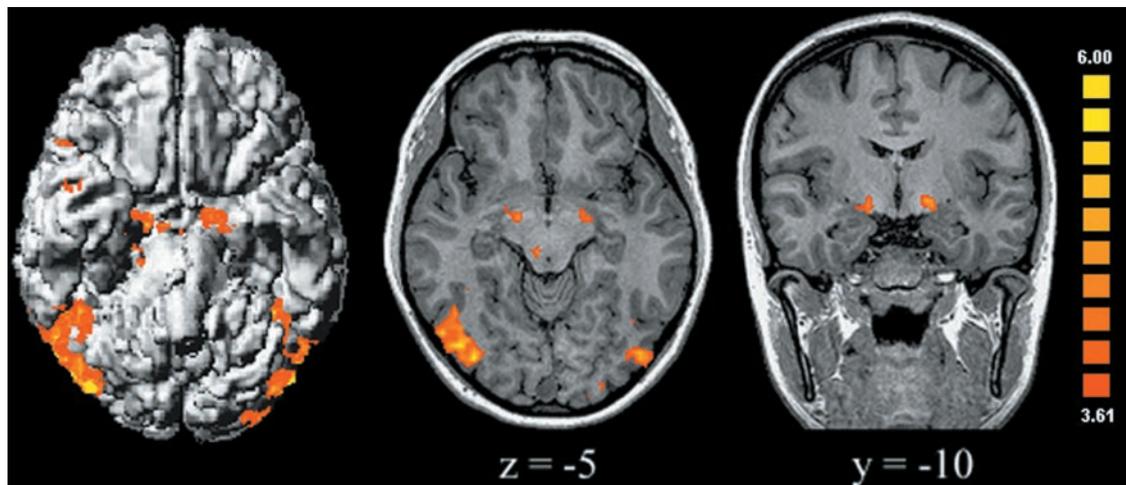
## RESULTS

### Behavioral data

Analysis of behavioral data validated all experimental conditions as members of the intended categories (Table 1). The moral content of the moral pictures was rated as higher than that of any other picture set ( $p < 0.00001$ ). This result did not change after arousal level and emotional valence were entered as covariates ( $F_{(5,856)} = 38.71$ ;  $p < 0.001$ ), confirming that the moral content could be treated as a valid independent category. Moral ratings were equivalent for unpleasant, pleasant, interesting, and neutral conditions ( $p > 0.22$ ). Stimuli belonging to the moral and unpleasant conditions were rated as having more negative emotional valence (i.e., unpleasantness) than all other picture sets ( $p < 0.00001$ ). Furthermore, unpleasant stimuli were rated as more unpleasant than the moral ones ( $p < 0.00001$ ). As expected, the pleasant pictures were judged to have the most positive valence. Moral, pleasant, unpleasant, and interesting conditions were associated with equivalent degrees of arousal ( $p > 0.24$ ) and were significantly more arousing than the neutral condition ( $p < 0.0001$ ). This latter finding is important, because the level of arousal appears to exert an effect on brain activation that is partially independent of emotional valence (Lane et al., 1999).

### Common effects of moral and unpleasant conditions

We first used conjunction analysis to explore the common effects of moral pictures and nonmoral unpleasant pictures relative to neutral pictures (Price and Friston, 1997). Common effects were observed in a network of limbic/paralimbic, subcortical, and cortical regions that included the extended amygdala and upper midbrain bilaterally, periaqueductal gray matter, right thalamus



**Figure 1.** Areas activated by both moral unpleasant and basic unpleasant conditions relative to the neutral condition (assessed with conjunction analysis) are shown. Activations were observed in the amygdala, upper midbrain, right thalamus, superior colliculus, extrastriate visual cortex, temporal neocortex, and insula (uncorrected  $p < 0.0001$ ; adjacent voxels in the cluster with  $t > 3.61$ ,  $p < 0.003$ ; minimum cluster volume of  $100 \text{ mm}^3$ ). In transversal and coronal slices, the right side of the brain corresponds to the left side of the figure.

and superior colliculus, right insula/inferior frontal gyrus [Brodmann area (BA) 44/45], right anterior temporal cortex (BA 21/38), bilateral posterior temporal–occipital cortex (BA 22/37/19), and right intraparietal sulcus (BA 7) (Fig. 1). These results replicate previous functional imaging findings on the neural substrates of unpleasant emotion (Lane et al., 1999; Damasio et al., 2000); they also demonstrate that moral and basic unpleasant emotions share common neural substrates. Accordingly, impaired emotional behavior has been consistently observed after damage to the thalamus, amygdala, and insula (Fukatsu et al., 1997; Adolphs, 1999; Calder et al., 2000).

#### Specific effects of processing moral stimuli

Our main goal was to identify the brain regions that are involved in moral processing, which was assessed by directly contrasting the moral and unpleasant conditions. The most striking findings were the increased activation of the right medial OFC and the medial frontal gyrus (MedFG) and the cortex surrounding the right posterior superior temporal sulcus (STS) elicited by the moral in comparison with the nonmoral unpleasant stimuli (Table 2, Fig. 2). Indeed, viewing moral scenes consistently activated these regions differentially, as revealed by contrasting the moral condition with each of the neutral (Fig. 3), pleasant, and interesting conditions. These findings support a specific role for this network in processing moral stimuli, because the effects of emotional valence and visual arousal were specifically controlled.

#### Specific effects of nonmoral unpleasant stimuli

To probe the brain regions more strongly activated by nonmoral unpleasant than by moral pictures, the unpleasant condition was compared with the moral condition. Nonmoral unpleasant stimuli (which included mangled faces, a snake, and a dirty toilet) were evocative of aversive emotions such as disgust and fear. The only brain regions significantly more active for this contrast were the right middle frontal gyrus and the right anterior insula (Table 2, Fig. 4). This latter finding concurs with lesion and functional imaging studies, which suggest a critical role of the anterior insula in disgust processing (Phillips et al., 1997; Calder et al., 2000). Interestingly, no amygdala activation was observed, suggesting that it was equally activated in unpleasant and moral conditions.

This could be explained by the presence of fearful stimuli in both conditions. As expected, when the basic unpleasant condition was contrasted with the neutral one, additional activations were observed in the amygdala, thalamus, and extrastriate visual cortex, brain regions shown in previous studies to be activated by processing visual stimuli associated with unpleasant emotions (Lane et al., 1999) (Fig. 5). No significant OFC/MedFG activations were observed for these contrasts.

#### Differential effects of faces and moral stimuli on brain activation

To rule out the possibility that activations were attributable solely to the presence of faces in the moral pictures, we also analyzed the data by creating regressors that represented (1) pictures displaying people and faces from nonmoral conditions and (2) moral pictures. These regressors were balanced, each containing an equal number of time points. Regions more strongly activated by faces compared with moral pictures were found in the fusiform gyri (BA 19), right lingual/inferior occipital gyrus (BA 18), and right intraparietal sulcus (BA 7/19). Regions more strongly activated by moral pictures compared with faces included the medial OFC/MedFG (BA 10/11), inferior frontal gyrus (BA 45), anterior temporal cortex (BA 21/38), posterior STS (BA 21/39), and both amygdala. These results demonstrate that the medial OFC/MedFG and posterior STS are centrally involved in processing pictures with high moral content, in contrast to the confirmed role of the fusiform gyrus in face processing (Kanwisher et al., 1997). The consistency of these activations with regard to previous studies and to our predictions confirms that subjects were effectively attending to the stimuli.

#### Condition-dependent changes in the connectivity of the medial OFC

To investigate further the brain circuits involved in moral processing, we studied condition-dependent changes in “functional connectivity” to test the hypothesis that a specific network of regions mediates the processing of stimuli with high socio-moral content. This approach identifies changes in the differential coupling among brain regions (i.e., it highlights the contribution of one area to the signal measured in other areas as a function of the

**Table 2. Anatomical locations and Talairach coordinates of activations**

Brain region, Brodmann area	Center Talairach coordinates		
	x	y	z
<b>Moral versus unpleasant</b>			
Right MedFG, 10	6	57	1
Right medial OFC, 11	17	40	−4
Right posterior STS, 21/39	41	−47	22
Left posterior middle temporal gyrus, 22/19	44	−64	9
Left lateral occipital gyrus, 18/19	−42	−69	8
<b>Unpleasant versus moral</b>			
Right middle frontal gyrus, 45/46	28	38	24
Right anterior insula	28	26	11
<b>Moral versus neutral</b>			
Right MedFG (frontopolar), 10	6	55	−1
Right MedFG, 10	8	41	17
Right medial OFC, 11	17	42	−7
Right/left amygdala/midbrain	17	−11	−6
	−16	−10	−5
Right anterior STS, 21/38	47	5	−12
Right/left posterior middle temporal gyri, 22/19	45	−63	5
	−45	−70	6
Right/left fusiform gyri, 19	40	−45	−14
	−38	−44	−17
Right precuneus, 7	10	−55	43
Left inferior occipital gyrus, 18	−40	−77	−18
<b>Unpleasant versus neutral</b>			
Right inferior and middle frontal gyri, 45/46	41	14	18
Right/left amygdala/midbrain	15	−9	−5
	−13	−8	−4
Right thalamus/superior colliculus	8	−15	−1
Right anterior STS, 21/38	43	3	−15
Right anterior insula	30	27	7
Right lateral occipital/fusiform gyri, 19	43	−64	−2
Left lateral occipital gyrus, 18/19	−37	−77	−3
Left fusiform gyrus, 19	−40	−55	−17
Left cerebellum	−6	−78	−23
	−20	−75	−27
Periaqueductal gray matter	−2	−28	−2
Right intraparietal sulcus, 7	27	−80	21

experimental condition) (Friston et al., 1997). We tested whether there were condition-specific changes in the functional connectivity of the right medial OFC, which was the most consistent prefrontal activation related to moral stimuli, with all other voxels in the brain. When the moral condition was compared with the unpleasant condition, there was a highly specific and statistically significant increase in connectivity between the medial OFC and the STS, the precuneus, and the same regions of the MedFG described above (Fig. 6). The opposite comparison, unpleasant versus moral, did not reveal increases in connectivity between the medial OFC and the STS or other sectors of the prefrontal cortex. Instead, the medial OFC increased its connectivity with the extrastriate visual cortex. These results were reliable despite the use of a relatively long TR of 5 sec, which could have decreased our ability to gauge functional connectivity changes.

## DISCUSSION

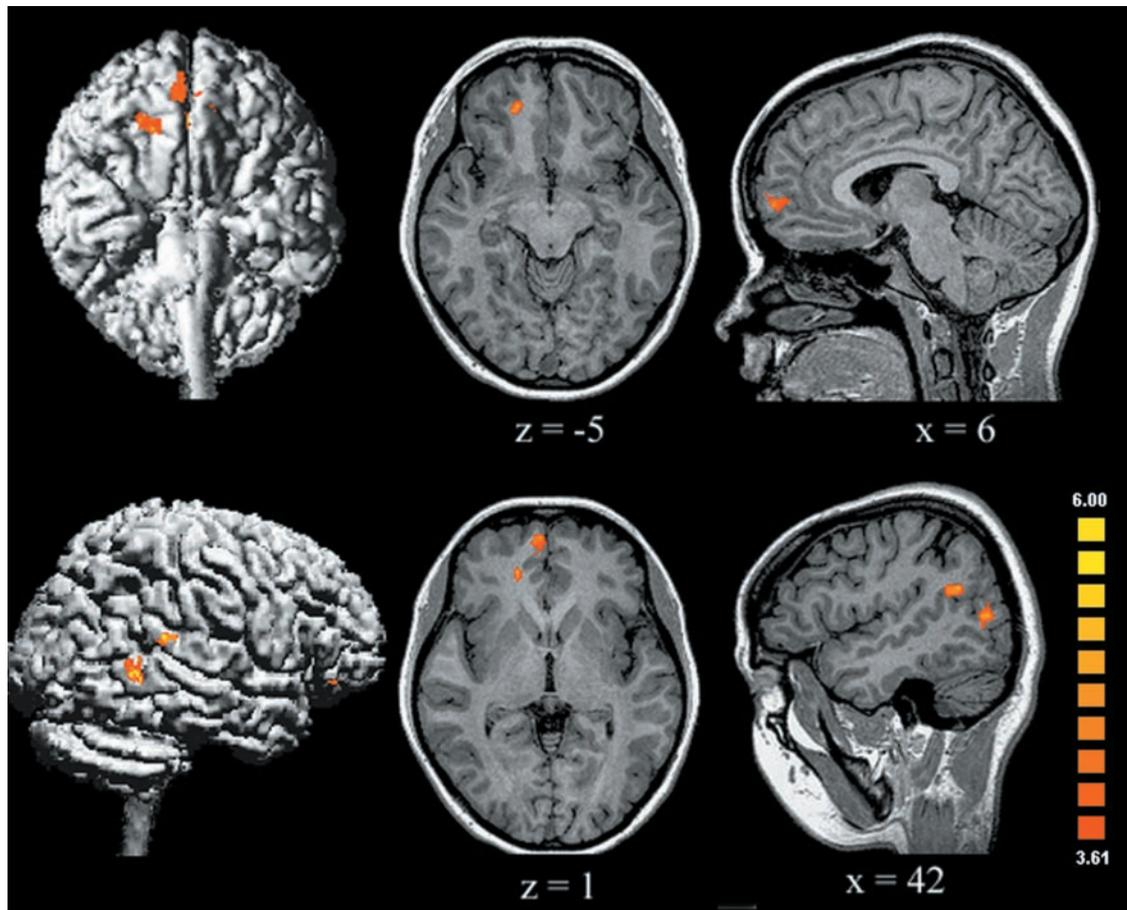
Our results show that viewing moral and nonmoral unpleasant visual stimuli activates a common network of brain areas that includes the amygdala, insula, thalamus, and upper midbrain. The medial OFC/MedFG and posterior STS were also recruited for processing emotionally charged moral stimuli. Behavioral results confirmed that the effects related to moral stimuli cannot be explained on the basis of the levels of emotional valence or visual arousal alone.

Moral emotions differ from basic emotions in that they are intrinsically interpersonal. Therefore, it was expected that moral emotions would share common neural substrates with tasks that invoke social schemas and behaviors, as well as inferences about the mental states of others. Imaging studies have implicated the ventral and medial prefrontal cortex in theory-of-mind tasks (Fletcher et al., 1995) and in the generation of emotional plans (Partiot et al., 1995). Brain-lesion studies have linked these regions to the development and maturation of moral, social, and emotional behavior (Eslinger et al., 1992; Anderson et al., 1999). The lateral superior temporal cortex, activated here when viewing moral pictures, plays an important role in the perception of social signs (Adolphs, 1999; Haxby et al., 2000). The present findings of preferential STS activation associated with moral stimuli and fusiform activation in response to face stimuli suggest that the STS may be involved in more complex social cognition mechanisms, whereas the fusiform gyrus might be more specifically linked to the processing of facial features per se (Kanwisher et al., 1997; Adolphs, 1999; Haxby et al., 2000). Unpublished observations from our group show that similar regions of the medial prefrontal cortex and STS are engaged when normal individuals are presented with statements evocative of moral emotions, compared with neutral statements and statements evocative of unpleasant, but nonmoral emotions. The stronger activation of the medial OFC and STS in response to both visual and lexical moral stimuli suggests that this network may be involved in moral processing regardless of the physical attributes of the stimuli.

Moral processing has been suggested to differ between men and women (Gilligan, 1993). Our study was not specifically designed to address gender effects on brain activation; however, visual inspection of the data show that the amygdala, STS, and medial OFC/MedFG were similarly activated in male and female subjects in response to moral and nonmoral stimuli. This indicates that both genders contributed to the findings reported above. Future imaging studies should address this relevant issue more directly in larger subject groups.

### OFC and automatic emotional processing

The OFC is an important mediator of automatic responses that guide motivated behavior (Damasio et al., 2000) and of implicit social-emotional appraisals (Damasio, 1994; Rolls et al., 1994; Milne and Grafman, 2001). The ventral and medial portions of the prefrontal cortex have been consistently activated in functional imaging studies of decision making related to reward and punishment (O'Doherty et al., 2001) and empathic and moral judgments (Farrow et al., 2001; Greene et al., 2001; Moll et al., 2001). Importantly, these studies have required explicit responses from subjects. In contrast, our task engaged subjects not as agents, but as observers, because no overt responses or explicit executive operations were required. Nevertheless, viewing pictures with moral content specifically activated the medial OFC/MedFG. These findings are consistent with the hypothesis that the OFC is involved in implicit executing behaviors that rely on the evalua-



**Figure 2.** The areas shown are those activated by moral versus unpleasant conditions; these areas included the medial OFC/MedFG, superior temporal sulcus, and posterior middle temporal gyrus of the right hemisphere (uncorrected  $p < 0.0001$ ; adjacent voxels in the cluster with  $t > 3.61$ ,  $p < 0.003$ ; minimum cluster volume of  $100 \text{ mm}^3$ ).



**Figure 3.** Areas activated by moral versus neutral conditions are shown. In addition to the prefrontal and temporal lobe activations, there was also activation of the amygdala bilaterally, the midbrain, extensive areas of the temporo-occipital cortex, and the precuneus (uncorrected  $p < 0.0001$ ; adjacent voxels in the cluster with  $t > 3.61$ ,  $p < 0.003$ ; minimum cluster volume of  $100 \text{ mm}^3$ ).

tion of reinforcement associations between environmental stimuli (Rolls et al., 1994). The medial OFC/MedFG, together with the STS and limbic-subcortical regions, may play an important role in the rapid and automatic detection of social-emotional events and in the induction of the cognitive, emotional, and motivational dispositions that have been proposed by modern theories as the

bases of moral sensitivity and behavior (Gilligan, 1993; Rozin et al., 1999; Greene et al., 2001; Haidt, 2002).

#### **Moral emotion, moral judgment, and moral behavior**

Patients who sustain damage to the medial OFC may have the ability to make moral and social judgments in laboratory settings



Figure 4. Areas activated by basic unpleasant versus moral stimuli were observed only in the right anterior insula and adjacent dorsolateral prefrontal cortex (uncorrected  $p < 0.0001$ ; adjacent voxels in the cluster with  $t > 3.61$ ,  $p < 0.003$ ; minimum cluster volume of  $100 \text{ mm}^3$ ).



Figure 5. When unpleasant stimuli were compared with the neutral stimuli, activation was observed in both the amygdala and the midbrain, as well as in the extrastriate visual cortex, in addition to the right insula and dorsolateral prefrontal cortex (uncorrected  $p < 0.0001$ ; adjacent voxels in the cluster with  $t > 3.61$ ,  $p < 0.003$ ; minimum cluster volume of  $100 \text{ mm}^3$ ).



Figure 6. Brain regions showing an increase in functional connectivity with the medial OFC in the moral condition relative to the unpleasant condition. For this analysis, the mean signal time courses of the right medial OFC cluster (center coordinate,  $x = 17$ ,  $y = 42$ ,  $z = -7$ ) were obtained from each subject (from the same cluster coordinates). Linear regression detected any voxel in the brain that increased its coupling with the signal of the right OFC. The statistical threshold was set at  $p < 0.002$  (uncorrected), adjacent voxels in the cluster with  $p < 0.005$  ( $t > 2.8$ ), and a minimum cluster volume of  $50 \text{ mm}^3$ . There was a marked increase in coupling of the right medial OFC with the MedFG, superior temporal sulcus, and precuneus.

even when their moral behavior in real-life situations is abnormal (Eslinger and Damasio, 1985; Saver and Damasio, 1991). We suggest that an impaired ability to automatically and rapidly process moral emotions in response to signs of moral violations may be a critical mechanism underlying this dissociation. This hypothesis is in line with psychological studies showing that moral behavior covaries with moral emotion more than with moral judgment (Haidt, 2001). Preston and de Waal (2002) have suggested that the automatic nature of this process is a pervasive feature of social–emotional perception–action mechanisms.

We reported previously that brain activation patterns are elicited by an explicit moral reasoning task (Moll et al., 2001). In that study, normal individuals were asked to judge statements with moral content (e.g., “we break the law when necessary”) and factual statements devoid of moral connotations (e.g., “stones are lighter than water”) as right or wrong. Group results showed activations predominantly in the frontopolar cortex, more rostrally than those reported here. Similar results have been reported by other investigators (Greene et al., 2001). Together, these results suggest that, in comparison with the automatic processing performed by the OFC in moral emotions, explicit moral reasoning may engage additional prefrontal regions (Eslinger et al., 2002).

In summary, based on our present and previous results, we suggest that the medial OFC/MedFG and the STS are critical elements of a cortical–limbic network that enables humans to link emotional experience to moral appraisals.

## REFERENCES

- Adolphs R (1999) Social cognition and the human brain. *Trends Cogn Sci* 3:469–479.
- 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.
- Blair RJR (1995) A cognitive developmental approach to morality: investigating the psychopath. *Cognition* 57:1–29.
- Blair RJR, Sellars C, Strickland I, Clark F, Williams AO, Smith M, Jones L (1995) Emotion attributions in the psychopath. *Pers Individ Dif* 19:431–437.
- Calder AJ, Keane J, Manes F, Antoun N, Young AW (2000) Impaired recognition and experience of disgust following brain injury. *Nat Neurosci* 3:1077–1078.
- Cohen O, River Y, Abramsky O (1999) Seizures induced by frustration and despair due to unresolved moral and political issues: a rare cause of reflex epilepsy. *J Neurol Sci* 162:94–96.
- Damasio AR (1994) *Descartes' error: emotion, reason, and the human brain*. New York: Avon.
- Damasio AR, Grabowski TJ, Bechara A, Damasio H, Ponto LL, Parvizi J, Hichwa RD (2000) Subcortical and cortical brain activity during the feeling of self-generated emotions. *Nat Neurosci* 3:1049–1056.
- de Waal F (1996) *Good natured: the origins of right and wrong in humans and other animals*. Cambridge, MA: Harvard UP.
- Eslinger PJ, Damasio AR (1985) Severe disturbance of higher cognition after bilateral frontal lobe ablation: patient EVR. *Neurology* 35:1731–1741.
- Eslinger PJ, Grattan LM, Damasio H, Damasio AR (1992) Developmental consequences of childhood frontal lobe damage. *Arch Neurol* 49:764–769.
- Eslinger PJ, Moll J, Oliveira-Souza R (2002) Emotional and cognitive processing in empathy and moral behavior. *Behav Brain Sci*, in press.
- Farrow TFD, Zheng Y, Wilkinson ID, Spence SA, Deakin JF, Tarriner N, Griffiths PD, Woodruff PW (2001) Investigating the functional anatomy of empathy and forgiveness. *NeuroReport* 12:2433–2438.
- Fletcher PC, Happé F, Frith U, Baker SC, Dolan RJ, Frackowiak RS, Frith CD (1995) Other minds in the brain: a functional imaging study of “theory of mind” in story comprehension. *Cognition* 57:109–128.
- Forman SD, Cohen JD, Fitzgerald M, Eddy WF, Mintun MA, Noll DC (1995) Improved assessment of significant activation in functional magnetic resonance imaging (fMRI): use of a cluster-size threshold. *Magn Reson Med* 33:636–647.
- Friston KJ, Holmes AP, Poline JB, Grasby PJ, Williams SC, Frackowiak RS, Turner P (1995) Analysis of fMRI time-series revisited. *NeuroImage* 2:45–53.
- Friston KJ, Buechel C, Fink GR, Morris J, Rolls ET, Dolan RJ (1997) Psychophysiological and modulatory interactions in neuroimaging. *NeuroImage* 6:218–229.
- Fukatsu R, Fujii T, Yamadori A, Nagasawa H, Sakurai Y (1997) Persisting childish behavior after bilateral thalamic infarcts. *Eur Neurol* 37:230–235.
- Gilligan C (1993) *In a different voice*, Ed 3. Cambridge, MA: Harvard UP.
- Greene JD, Sommerville RB, Nystrom LE, Darley JM, Cohen JD (2001) An fMRI investigation of emotional engagement in moral judgment. *Science* 293:2105–2108.
- Haidt J (2001) The emotional dog and its rational tail: a social intuitionist approach to moral judgment. *Psychol Rev* 108:814–834.
- Haidt J (2002) The moral emotions. In: *Handbook of affective sciences* (Davidson RJ, Scherer K, Goldsmith HH, eds), Oxford: Oxford UP, in press.
- Haxby JV, Hoffman EA, Gobbini MI (2000) The distributed human neural system for face perception. *Trends Cogn Sci* 4:223–233.
- Kanwisher N, McDermott J, Chun MM (1997) The fusiform face area: a module in human extrastriate cortex specialized for face perception. *J Neurosci* 17:4302–4311.
- Lane R, Chua P, Dolan R (1999) Common effects of emotional valence, arousal and attention on neural activation during visual processing of pictures. *Neuropsychologia* 37:989–997.
- Lang PJ, Bradley MM, Cuthbert BN (1995) *International Affective Picture System (IAPS)*. Bethesda, MD: National Institute of Mental Health Center for the Study of Emotion and Attention.
- Milne E, Grafman J (2001) Ventromedial prefrontal cortex lesions in humans eliminate implicit gender stereotyping. *J Neurosci* 21:RC150:1–6.
- 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.
- O’Doherty J, Kringelbach ML, Rolls ET, Hornak J, Andrews C (2001) Abstract reward and punishment representations in the human orbitofrontal cortex. *Nat Neurosci* 4:95–102.
- Partiot A, Grafman J, Sadato N, Wachs J, Hallett M (1995) Brain activation during the generation of non-emotional and emotional plans. *NeuroReport* 10:1269–1272.
- 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.
- Preston SD, de Waal FBM (2002) Empathy: its ultimate and proximate bases. *Behav Brain Sci*, in press.
- Price BH, Daffner KR, Stowe RM, Mesulam MM (1990) The compartmental learning disabilities of early frontal lobe damage. *Brain* 113:1383–1393.
- Price CJ, Friston KJ (1997) Cognitive conjunction: a new approach to brain activation experiments. *NeuroImage* 5:261–270.
- Rolls ET, Hornak J, Wade D, McGrath J (1994) Emotion-related learning in patients with social and emotional changes associated with frontal lobe damage. *J Neurol Neurosurg Psychiatry* 57:1518–1524.
- Rozin P, Loewry 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.
- Saver JL, Damasio AR (1991) Preserved access and processing of social knowledge in a patient with acquired sociopathy due to ventromedial frontal damage. *Neuropsychologia* 29:1241–1249.
- Talairach J, Tournoux P (1988) *Co-planar stereotaxic atlas of the human brain*. New York: Thieme Medical.