Functional Networks in Emotional Moral and Nonmoral Social Judgments

Jorge Moll,* Ricardo de Oliveira-Souza,* Ivanei E. Bramati,* and Jordan Grafman†

*Neuroimaging and Behavioral Neurology Group (GNNC), Hospitals D'Or and LABS, Rio de Janeiro, RJ, Brazil; and
†Cognitive Neuroscience Section, National Institute of Neurological Disorders and Stroke, Bethesda, Maryland

Reading daily newspaper articles often evokes opinions and social judgments about the characters and stories. Social and moral judgments rely on the proper functioning of neural circuits concerned with complex cognitive and emotional processes. To examine whether dissociable neural systems mediate emotionally charged moral and nonmoral social judgments, we used a visual sentence verification task in conjunction with functional magnetic resonance imaging (fMRI). We found that a network comprising the medial orbitofrontal cortex, the temporal pole and the superior temporal sulcus of the left hemisphere was specifically activated by moral judgments. In contrast, judgment of emotionally evocative, but non-moral statements activated the left amygdala, lingual gyri, and the lateral orbital gyrus. These findings provide new evidence that the orbitofrontal cortex has dedicated subregions specialized in processing specific forms of social behavior. © 2002 Elsevier Science (USA)

Key Words: moral judgment; fMRI; orbitofrontal; acquired sociopathy; frontal lobes.

INTRODUCTION

Humans routinely judge the social behavior of others (Brothers, 1997). These judgments can be made on the basis of moral beliefs or can be just an interpretation of an actor's intention and the several possible outcomes of that intention (Colby, 1990; Haidt, in press). Such complex social judgments rely on the proper functioning of brain networks dedicated to processing stimuli endowed with social and emotional significance (Eisenberg, 1995; Damasio, 2000). The integration of cognitive evaluation with emotional bias allows humans to more confidently judge the social consequences of other people's actions (Fletcher *et al.,* 1995; Frith and Frith, 1999). Damage to these systems may lead to distinct social behavior abnormalities (Eslinger and Damasio, 1985; Adolphs, 1999). The neural systems underlying these social abilities include the orbitofrontal cortex (OFC), the temporal neocortex and the amygdala

(Brothers, 1997; Adolphs *et al.,* 1998; Adolphs, 1999). While numerous functional imaging studies in humans have demonstrated cortical and limbic activation when subjects visually process social (see Adolphs, 1999, for a review) or emotional stimuli (Reiman *et al.,* 1997; Lane *et al.,* 1999), only a few have examined complex social cognition (Hoffman and Haxby, 2000; Golby *et al.,* 2001). Investigations of patients with autism (Frith and Frith, 1999) and psychopaths (Blair, 1995) suggest that different aspects of social cognition can be selectively impaired. For example, although autistic individuals have generally impaired social behavior (Frith and Frith, 1995), psychopaths are impaired on a subset of social behaviors that require moral appraisals (Blair, 1995).

The neurological and neuropsychiatric literature is also informative in regard to the organization of the human social brain. There is increasing evidence that primary psychopaths comprise a particularly severe subgroup of antisocial individuals whose deviant behaviors are related to structural and functional brain abnormalities (Raine *et al.,* 2000; Kiehl *et al.,* 2001). Acquired brain damage in previously normal individuals can also lead to a set of related clinical syndromes reminiscent of primary psychopathy (Tranel, 1994). The antisocial behavior of these patients may result from sheer impulsiveness and goal neglect to recurrent flagrant criminal and evil actions that represent a bizarre change from their premorbid personality styles (Brower and Price, 2002). Like primary psychopaths, such "acquired sociopaths" often retain the ability to tell right from wrong and to articulate sound statements on morality and social appropriateness, that stands in sharp contrast to their behavior in real life. The damage in such cases falls within an extended area that encompasses the polar and mediobasal divisions of the frontal lobes (Kandel and Freed, 1989), the temporal poles (Miller *et al.,* 1999) and several basal forebrain structures that are interconnected by the medial forebrain bundle and extend from the ventromedial hypothalamus caudally to the amygdala and

septal area rostrally (Adolphs *et al.,* 1998; Flynn *et al.,* 1988; Gorman and Cummings, 1992).

In previous reports (Oliveira-Souza and Moll, 2000; Moll *et al.,* 2001), we addressed the brain networks involved in the processing of sentences with or without moral content using fMRI in normal subjects. A selective network of brain regions was more active when subjects judged moral as opposed to factual statements. These regions included the ventral-anterior and medial sectors of the prefrontal cortex (frontopolar and medial frontal gyri), right anterior temporal cortex, left angular gyrus, and globus pallidus. Moreover, while the emotional valence of stimuli seemed to be directly related to the activations in right anterior temporal lobe and subcortical nuclei, it played only an ancillary role in the prefrontal cortex activation. Despite such compelling evidence, since our previous study design did not include emotional stimuli as an independent, nonmoral, experimental condition, it is difficult to conclude which brain regions were distinctively recruited by emotion processing as opposed to moral judgment. Another recent fMRI study (Greene *et al.,* 2001) reported similar activation of the anterior prefrontal cortex in response to complex moral judgments. This study did not include a nonmoral emotional condition either and, thus, the reported activations could have been induced by emotional processing. Besides, the increased attentional and decision-making demands of the moral-emotional condition might be associated with slower reaction times, which could by themselves have led to increased prefrontal activation (de Zubicaray *et al.,* 2001). In our previous reports (Oliveira-Souza and Moll, 2000; Moll *et al.,* 2001), we employed a measure of judgment difficulty and found equivalent judgment difficulty for moral and nonmoral stimuli, favoring the view that cognitive effort was not decisively involved in the brain activation patterns. Notwithstanding these observations, the issue of which brain regions are specifically engaged when normal subjects judge moral and nonmoral, emotionally charged situations remains unsettled.

Here we employed functional MRI to address the role of emotional valence and moral content in social judgments. For this purpose, normal subjects were requested to judge statements pertaining to three main conditions: emotionally unpleasant statements without moral connotations, emotionally unpleasant statements with moral connotations, and emotionally neutral statements. Scrambled statements were included as a baseline condition. By including an experimental condition evocative of emotions devoid of moral connotations, we intended to explore the differential effect of emotional valence and moral judgment on brain activation. An additional feature of the present study was that statements pertaining to all experimental conditions described human actions unfolding in social scenarios. Based on the evidence that different kinds of

social abilities and emotional processing may be dissociated in cases of brain damage (Eslinger and Damasio, 1985; Blair, 1995), we hypothesized that the judgment of unpleasant-moral and unpleasant–nonmoral social situations would lead to distinct patterns of corticolimbic activation. More specifically, we expected that the anterior sectors of the prefrontal cortex, as shown by our previous study (Moll *et al.,* 2001), would be preferentially activated by moral judgments when compared with judgments of emotional situations without moral connotations. We also predicted that the judgment of unpleasant scenarios devoid of moral content would activate brain regions that mediate basic unpleasant emotions, such as the amygdala.

MATERIALS AND METHODS

Subjects. Seven healthy right-handed adults (three males, mean age of 30.3 ± 4.7 years, 14.9 ± 2.0 years of education), with no history of neuropsychiatric disorders, participated in the fMRI study. An independent group of seven subjects whose age, level of education and gender distribution did not differ from the fMRI subjects (age: $t = 1.76$, $P = 0.11$; education: $t = 1.13$, $P = 0.28$; gender: $x^2 = 1.40$, $P = 0.56$), was recruited for a supplemental behavioral study. All subjects gave written informed consent and did not receive financial compensation. The study was conducted in the Hospital Barra D'Or and approved by the hospital's Institutional Review Board and Ethics Committee.

Stimuli and task. Subjects were asked to read short statements that were visually presented through LCD goggles (Resonance Technologies, Inc., CA) and to covertly judge them as being either right or wrong. We chose the words "certo" and "errado," the Portuguese equivalents of "right" and "wrong" because, as in English, they allow moral, factual, and structural connotations. Subjects understood that they were intended to use right or wrong attributions in a broadly defined manner. By limiting the output of their cognitive operations to only two categories ("right" or "wrong"), we expected to force each participant to make standard decisions regardless of the specific content of each sentence.

Rationale behind moral and nonmoral judgments. The distinction between moral and nonmoral judgments relies on psychological constructs that can be objectively assessed and have been extensively validated across different cultures (Snarey, 1985; Colby, 1990). A further distinction between nonmoral social norm violations and moral–social violations has also been supported by empirical evidence. This distinction forms the basis of the concept of conventional and moral transgressions, whose psychological bases have been worked out in normal individuals, psychopaths and in patients with autism, in both adults and children (Blair, 1995, 1996; Fisher and Blair, 1998).

TABLE 1

Sample Statements

Nonmoral neutral (NTR)
  He never uses the seat belt.
  The elderly sleep more at night.
  Fat children should make a diet.
  The painter used his hand as a paintbrush.
  Judges use white uniforms.
Nonmoral unpleasant (NM)
  He licked the dirty toilet.
  The elderly are used to eating living toads.
  Pregnant women often throw up.
  People don't have tatoos inside their eyeballs.
  Judges often eat rotten food.
Moral (M)
  He shot the victim to death.
  The elderly are useless.
  Criminals should go to jail.
  The father never treated his son as a slave.
  The judge condemned the innocent man.
Scrambled (SCR)
  Sons push use eat work.
  Kick like poor rain old have.
  Life turn of shoes drink was brother.
  Your drive building must meat.
  Day daughter less ground parents loose.

Experimental conditions. Three main conditions included structurally similar neutral (NTR), moral (M), and nonmoral (NM) statements (see Table 1 for a sample of the statements). NTR statements described relatively unemotional situations, NM statements described emotionally aversive scenarios without moral content, while M statements described emotional situations designed to evoke moral attitudes and feelings. All statements described scenarios and people acting in different settings, and, thus, each of the NTR, NM, and M statements deployed explicit social contexts. An additional condition made up of scrambled statements (SCR) was added to serve as a low level baseline. Subjects covertly judged all the sentences in a categorical fashion as right or wrong. After fMRI scanning, the sentences were presented again, this time in a randomized, unblocked fashion, and subjects were asked to overtly judge them as right or wrong by referring to their impressions while in the scanner. Subjects also rated the degree of moral content and of the positive or negative emotional impact of each statement on 4-point Likert scales. They were instructed that moral content meant issues of values, rights, justice, responsibilities and principles regarding peace and care for others (Colby, 1990; Gilligan, 1993). Subjects were encouraged to make a short verbal commentary about each statement. These comments were scored by two judges who used a predefined lexicon of words that expressed basic or moral emotions. The identification of basic in contrast to moral emotions in these descriptions was based on concepts firmly grounded in a large body of empirical research (Haidt, in press). This pro-

cedure helped us ascertain that the statements employed in the moral condition suited our purposes as moral eliciting stimuli.

Each condition was composed of a set of 24 statements that were presented in blocks of three at a time. Stimuli were displayed for 5 s and were separated by 5 s of a blank screen. A pilot study showed that this presentation rate allowed subjects to comfortably read and judge each sentence. The eight blocks of each condition followed a fixed pseudorandomized order. A fixation cross-hair was displayed for 15 s between each block to allow a complete return of the BOLD signal to baseline. Scrambled sentences were generated by randomly reordering words sampled from the other conditions, being semantically meaningless and grammatically incorrect ("nonsense"). Subjects were not informed beforehand about the content of the sentences, yet they were asked to "try and get their meaning and to judge all of them, regardless of how weird they might sound." This was meant to encourage subjects to stick to the scrambled sentences and prevent their attention from wandering. To minimize the effects of planning and to prevent the subjects from reasoning about the future, the situations depicted in all statements referred to the present or past only.

Behavioral measures. To obtain a measure of task difficulty, seven subjects not participating in the fMRI experiment judged the blocked 24 NTR, M, NM, and SCR statements. These blocks were administered in a randomized order across subjects. The time to judge each block was recorded as a response time (RT) and taken as a proxy for task difficulty. Although collecting responses during fMRI data acquisition would be more desirable, MR compatible response devices were not available in our institution at the time this study was completed. Although we could have obtained RTs from the same subjects by repeating the task after fMRI scanning, we pondered that subjects attitudes to repeated stimuli would not necessarily reflect their naïve responses. For this reason, a different group of subjects matched for age, gender, education and cultural background was employed. This approach has been successfully adopted in previous functional imaging studies (Bottini *et al.,* 1994).

Ratings for moral content, emotional impact, and mean response times were assessed with analysis of variance. Pairwise comparisons of means were evaluated post hoc with the least significant difference test. The association between emotional impact and moral content in the M condition was assessed with the Phi coefficient. Computations were performed using *Statistica,* v. 5.5 (StatSoft, 1999).

fMRI procedures and data analysis. Anatomic (3D-GRE T1-weighted images, 1.25 mm) and functional data (BOLD-EPI imaging, TR/TE = 4980/66 ms, 128 × 128, FOV = 256 mm, thickness/gap = 5/0.25 mm, 16

TABLE 2

Anatomical Locations and Coordinates of Activations

Brain region, Brodmann area[a]	Center Talairach coordinates			Cluster P value	Cluster volume[b]
	X	Y	Z		
M vs NTR[c]					
L Medial OFC, 10/11[d]	−10	46	−12	0.0002	171
L Temporal pole, 38	−33	19	−23	0.001	50[e]
L STS, 21/22[d]	−47	−40	−2	0.0004	148
NM vs NTR[c]					
R Fusiform G, 19[f]	31	−66	−20	0.0004	139
R Inf Occ G, 19	28	−88	−22	0.001	100
R Lingual G, 17/18[f]	5	−72	0	0.000001	1750
Bilat Calcarine S, 17	−11	−88	2	0.000004	247
L Amygdala[f]	−20	−12	−6	0.000001	253
L Lingual G, 17/18[f]	−23	−64	−15	0.000001	1068
L Lateral OFC, 11/47[f]	−26	32	−1	0.001	108
L Fusiform G, 19[f]	−34	−62	−11	0.00003	206

[a] OFC, orbitofrontal cortex; STS, superior temporal sulcus; G, gyrus; S, sulcus; Inf, inferior; Occ, occipital; L, Left; R, right; Bilat, bilateral.
[b] Cluster volumes of at least 100 mm³.
[c] M, NTR, and NM correspond to moral, neutral, and nonmoral social conditions, respectively.
[d] Also activated in the M vs NM comparison.
[e] Cluster volume below 100 mm³.
[f] Also activated in the NM vs M comparison.

slices) were obtained with a 1.5 MRI scanner (Siemens Vision, Germany). Data acquisition was synchronized with stimulus presentation. Functional datasets were 3D motion-corrected. Slice time correction, temporal smoothing, linear trend removal and spatial normalization (3-D Gaussian kernel = 4 mm) were performed. Datasets were co-registered and Talairach transformed (Talairach and Tournoux, 1988). Activation maps were analyzed with statistical parametric methods (Friston *et al.,* 1995) contained in BrainVoyager v. 3.9 (Brain Innovation, Maastricht, The Netherlands). Regressors representing the experimental conditions were modeled with a hemodynamic response filter entered in a multiple regression analysis, using a fixed-effects model. Significance was assessed using a threshold corresponding to $P < 0.001$ (uncorrected) at the voxel level and a three-dimensional cluster extent threshold of 100 mm³ to protect against Type I errors associated with multiple statistical comparisons (Forman *et al.,* 1995). Smaller activated clusters (50–100 mm³) were reported when they fell in a region where an *a priori* hypothesis predicted activation (OFC and medial prefrontal cortex, amygdala and anterior temporal cortex). Levels of statistical significance for each activated cluster are reported in the Table 2. Results were overlaid on averaged anatomical data from all subjects and on partially inflated three-dimensional templates of a sample subject.

RESULTS

Behavioral. Moral content was rated significantly higher in the M (2.12 ± 1.41) in comparison to the NTR (0.07 ± 0.37) and NM (0.27 ± 0.82) conditions [$F(2,499) = 228.9$, $P < 0.00002$]. The M and NM statements were designed so that they had negative emotional valence. This was validated by subject ratings of those same statements at debriefing. The emotional valence of the NTR condition approached zero (−0.05 ± 0.98), whereas the net emotional valence for both the M (−2.02 ± 1.26) and NM (−0.88 ± 1.20) conditions was negative. The NM and M conditions differed significantly from the NTR condition in degree of emotional impact [$F(2,497) = 138.23$, $P < 0.0001$]. M sentences were rated as having a higher emotional impact than NM sentences ($P < 0.0001$). There was a significant relationship between moral content and emotionality of statements ($r = 0.44$, $p < 0.0001$), suggesting that moral judgment interacts with the perceived emotionality of the stimulus. Judgments in the M condition were described more frequently with moral terms (e.g., "pity" and "indignation") than with basic emotion words (123 × 53) in contrast to judgments in the NM condition, in which primary emotion words (e.g., "disgust") were used more often (136 × 1) (Phi = 0.70, $P < 0.0001$). This result indicated that the M sentences were effective in eliciting moral concerns. There was an overall difference in judgment times across trials ([$F(3,18) = 3.42$, $P < 0.04$]. Post hoc analyses showed that this difference was due to slower RT in the SCR condition (92.3 ± 33.5 s, $P < 0.03$). Mean RT did not statistically differ for the NTR (75.3 ± 20.7 s), M (73.7 ± 24.3 s), and NM (76.7 ± 25.3 s) conditions, suggesting that judging them recruited equivalent degrees of effort.
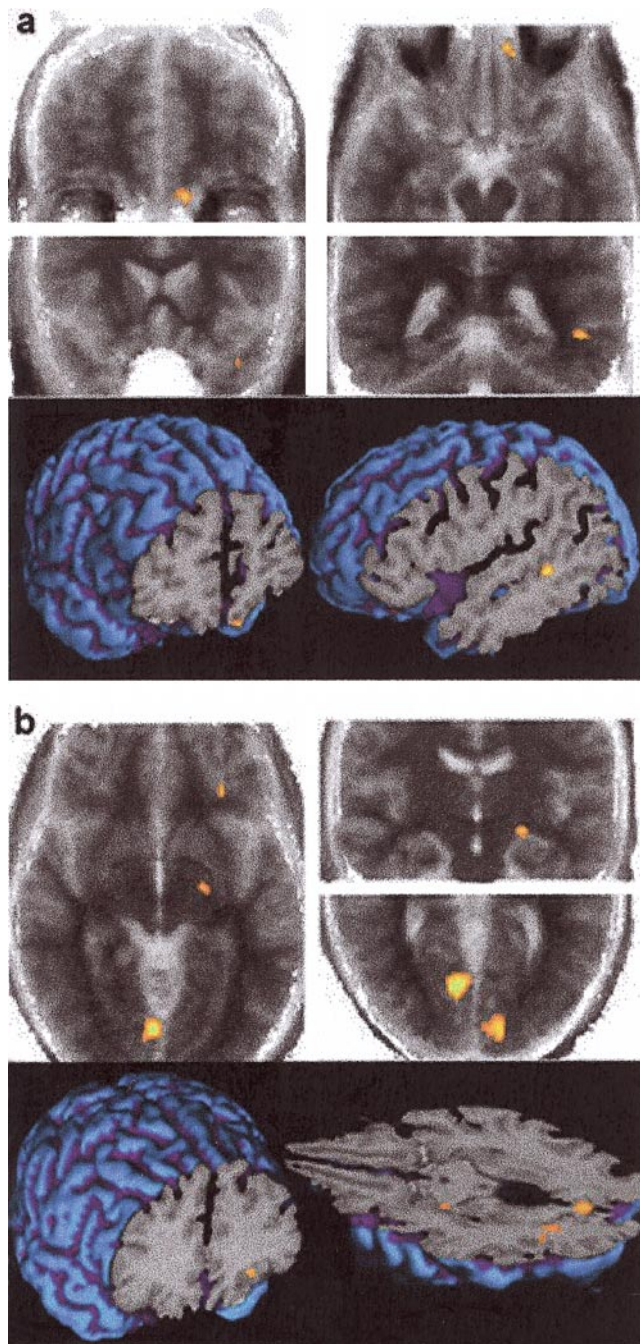
FIG. 1. Brain regions activated by emotionally evocative moral (M) and nonmoral (NM) judgments compared to neutral (NTR) ones. Activations were overlaid on sections through an averaged brain from all subjects with inverted grayscale and on 3-D renderings of a reference brain. (a) M vs NTR condition. Activated regions were in the left orbitofrontal cortex (OFC) and in the superior temporal sulcus and the left temporal pole. (b) NM vs NTR condition. Activated regions were in the left amygdala and lateral OFC, and bilaterally in the visual cortex.

Brain activation. Compared with the NTR condition, the M and NM conditions evoked distinct brain activation patterns (Figs. 1a and 1b and Table 2). The NM condition induced activation of the left amygdala and left lateral OFC as well as of several regions of the ventral visual cortex (lingual, inferior occipital and fusiform gyri). The M condition induced activation of the left medial OFC (gyrus rectus and medial orbital gyrus), left temporal pole, and the cortex of the superior temporal sulcus (STS), close to the angular gyrus.

When the M condition was compared to the NM condition, the same activations in the medial OFC (101 mm^3, $P < 0.001$) and STS (1063 mm^3, $P < 10^{-6}$) were seen. The left temporal pole activation was no longer observed. In the opposite NM vs M comparison, the left amygdala (242 mm^3, $P < 4 \times 10^{-6}$) and the lateral orbital gyrus (207 mm^3, $P < 3 \times 10^{-5}$) remained active, along with the visual cortex (lingual and fusiform gyri). Notably, with the exception of a limited sector of the STS that has been previously implicated in processing visual social cues (e.g., Hoffman and Haxby, 2000), other sectors of the temporal lobe were not activated when M and NM conditions were contrasted to each other. This is in agreement with the role of temporal lobe structures in semantic comprehension (Zahn *et al.,* 2000), and indicates that semantic processing was well matched in these conditions. Mean signal changes and standard errors from the medial and lateral OFC, amygdala, primary visual cortex and STS cortex in each experimental condition, averaged across subjects, are displayed in Fig. 2.

In order to investigate the effects of right vs wrong categorical judgments on brain activation, we computed Kendall's correlation coefficient between the percentage of BOLD signal increase in the left amygdala and medial OFC, and the number of statements judged as right vs wrong in the NM and M conditions, respectively. There was no relationship between the number of right vs. wrong judgments and the magnitude of the hemodynamic response of the amygdala ($T = -0.20$, $P > 0.62$) or medial OFC ($T = 0.09$, $P > 0.74$), suggesting that activation in these brain regions was independent of the frequency of outcome of the categorical right vs wrong judgments.

When the NTR condition was contrasted to the SCR condition, a pattern of brain activation similar to that described in studies of sentence processing was observed (Bottini *et al.,* 1994). Activated regions included the middle and posterior portions of the superior temporal gyrus and sulcus bilaterally (more extensively in the left hemisphere), the frontal opercula, the anterior cingulate and adjacent supplementary motor cortex, the thalamus and putamen, and additional foci in the anterior temporal lobes, fusiform gyrus, and cuneus (Fig. 3).

DISCUSSION

Our results provide new evidence that distinct neural networks are activated by different kinds of social
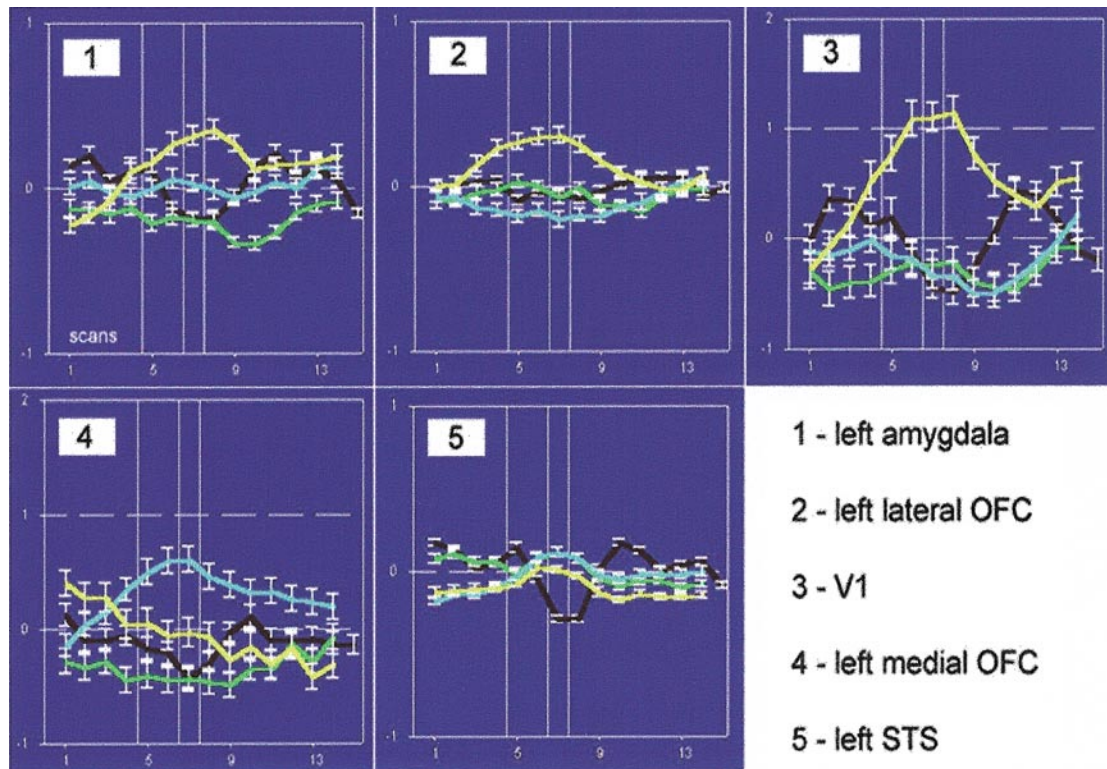
FIG. 2. Mean signal changes of the left medial and lateral OFC, left amygdala, primary visual cortex (V1), and the cortex of the left superior temporal sulcus (STS), obtained from averaged MR signal from all subjects. Curve colors correspond to experimental conditions as follows: yellow, unpleasant condition; light blue, moral condition; green, neutral condition; black, scrambled condition.

judgment. In particular, moral judgments associated with unpleasant emotions induced activation in the anterior aspect of the medial OFC, whereas nonmoral social judgments associated with unpleasant emotions induced lateral OFC and amygdala activation. The amygdala has a major role in processing emotionally arousing stimuli, both pleasant and aversive, and it

FIG. 3. Brain regions activated by the Neutral (NTR) as compared to Scrambled (SCR) condition (both temporal lobes and frontal opercula, supplementary motor area, anterior cingulate, basal ganglia, and thalamus).

has been suggested that it may help allocate resources to processing different kinds of biologically salient stimuli (Adolphs *et al.,* 1998; Golby *et al.,* 2001).

Activation of the left amygdala by a cognitive-emotional elicitation procedure (e.g., through language) implies that top-down mechanisms may activate this brain region under these conditions, in accord with recent functional imaging studies employing threatening words (Isenberg *et al.,* 1999) and the cognitive representation of fear (Phelps *et al.,* 2001). This finding is also consistent with the suggestion that the left amygdala, rather than the right, is more closely related to linguistic affective processes (Markowitsch, 1998; Phelps *et al.,* 2001). Contrary to our expectations, the amygdala was not activated in the moral judgment condition, even though those statements were rated as most emotionally evocative. A possible explanation is that the medial OFC, which is intimately linked to the processing of social rules and emotions related to moral processing, down-regulates the activity of the amygdala in certain circumstances (Baxter *et al.,* 2000). The amygdala is densely interconnected with the visual cortex, which is strongly activated by aversive pictures or words (Reiman *et al.,* 1997; Lane *et al.,* 1999). Thus, the activation of the visual cortex in the NM condition is not surprising, and could have resulted from modulatory effects from the amygdala (Morris *et al.,* 1998).
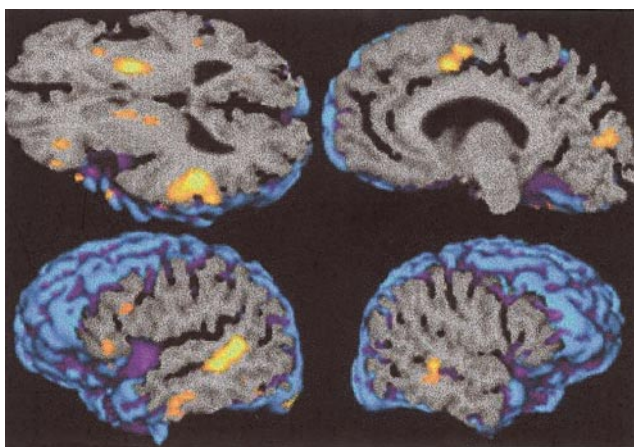
The amygdala is also massively interconnected with the OFC, especially with the caudal sector of its lateral subdivision (Baxter *et al.,* 2000; Öngur and Price, 2000), which was activated in the NM condition. This region is activated in abstract reward/punishment acquisition in humans (O'Doherty *et al.,* 2001) and, when damaged, is thought to impair social behavior (Anderson *et al.,* 1999). The medial OFC activation in the M condition is compatible with evidence showing that humans sustaining lesions in this region frequently present with social disinhibition, lack of empathy and increased levels of aggression (Grafman *et al.,* 1996; Anderson *et al.,* 1999; Pietrini *et al.,* 2000; Raine *et al.,* 2000). One explanation for these impairments is that patients are deficient in their ability to choose among behavioral alternatives based on inferences about positive or negative outcomes and changing reinforcement contingencies (Rolls *et al.,* 1994; Bechara *et al.,* 2000). The present findings suggest that the medial OFC may be even more critical for the integration of moral knowledge with the emotions that determine the reinforcing value of specific behavioral actions. It is also likely that the medial OFC, which receives projections from the STS region, integrates cues about the intentional and emotional states of others via signals from surface features of stimuli such as facial expression, body posture, and voice inflexions into decision-making (Adolphs, 1999; Hoffman and Haxby, 2000). The finding that the temporal pole was activated in the M vs NTR contrast, but not in the M vs NM or NM vs M comparisons, favors the view that it participates nonspecifically in both moral and nonmoral emotional processing. This result is consistent with the weaker activation of the right anterior temporal cortex when emotional impact was covaried in the design matrix in a previous study (Moll *et al.,* 2001).

As reported above, different sectors of the OFC were activated when moral or nonmoral social judgments were being contemplated. Since RTs were equivalent for the M and NM conditions, we believe that these different patterns of activation are not attributable to task difficulty or effort. The dissociable networks we have identified probably work in an integrated fashion during many social interactions that combine moral, social, and emotional demands. One limitation of our findings regards agency. The judgments employed in our study did not require reference to the subjects' own behavior. On evolutionary grounds, it is likely that agency, or role-taking (Ruby and Decety, 2001), critically influences social and emotional-based reasoning and behavior and can have a direct impact on the outcomes of goal-directed behavior. Examining the brain activation patterns related to the differential effects of agency and emotion on moral and factual judgments is a logical next step for researchers interested in the brain representation of morality-influenced behavior.

REFERENCES

Adolphs, R. 1999. Social cognition and the human brain. *Trends Cogn. Sci.* **3:** 469–479.

Adolphs, R., Tranel, D., and Damasio, A. R. 1998. The human amygdala in social judgment. *Nature* **393:** 470–474.

Anderson, S. W., Bechara, A., Damasio, H., Tranel, D., and Damasio, A. R. 1999. Impairment of social and moral behavior related to early damage in human prefrontal cortex. *Nature Neurosci.* **2:** 1032–1037.

Baxter, M. G., Parker, A., Lindner, C. C. C., Izquierdo, A. D., and Murray, A. E. 2000. Control of response selection by reinforcer value requires interaction of amygdala and orbital prefrontal cortex. *J. Neurosci.* **20:** 4311–4319.

Bechara, A., Tranel, D., and Damasio, H. 2000. Characterization of the decision-making efect of patients with ventromedial prefrontal cortex lesions. *Brain* **123:** 2189–2202.

Blair, R. J. R. 1995. A cognitive developmental approach to morality: Investigating the psychopath. *Cognition* **57:** 1–29.

Blair, R. J. R. 1996. Brief report: Morality in the autistic child. *J. Autism Dev. Disord.* **26:** 571–579.

Bottini, G., Corcoran, R., Sterzi, R., Paulesu, E., Schenone, P., Scarpa, P., Frackowiak, R. S., and Frith, C. D. 1994. The role of the right hemisphere in the interpretation of figurative aspects of language. A positron emission tomography activation study. *Brain* **117:** 1241–1253.

Brothers, L. 1997. *Friday's Footprint. How Society Shapes The Human Mind.* Oxford, New York.

Brower, M. C., and Price, B. H. 2002. Neuropsychiatry of frontal lobe dysfunction in violent criminal behavior: A critical review. *J. Neurol. Neurosurg. Psychiatry* **71:** 720–726.

Colby, A., Kohlberg, L., Speicher, B., Hewer, A., Candee, D., and Gibss, J. 1990. *The Measurement of Moral Judgment.* Cambridge, New York.

Damasio, A. R., Grabowski, T. J., Bechara, A., Damasio, H., Ponto, L. L. B., Parvizi, J., and Hichwa, R. D. 2000. Subcortical and cortical brain activity during the feeling of self-generated emotions. *Nature Neurosci.* **3:** 1049–1056.

de Zubicaray, G. I., Wilson, S. J., McMahon, K. L., and Muthiah, S. 2001. The semantic interference effect in the picture-word paradigm: An event-related fMRI study employing overt responses. *Hum. Brain Mapp.* **14:** 218–227.

Eisenberg, L. 1995. The social construction of the human brain. *Am. J. Psychiatry* **152:** 1563–1575.

Eslinger, P. J., and Damasio, A. R. 1985. Severe disturbance of higher cognition after bilateral frontal lobe ablation: Patient EVR. *Neurology* **35:** 1731–1741.

Fisher L., and Blair R. J. 1998. Cognitive impairment and its relationship to psychopathic tendencies in children with emotional and behavioral difficulties. *J. Abnorm. Child Psychol.* **26:** 511–519.

Fletcher, P. C., Happé, F., Frith, U., Baker, S. C., Dolan, R. J., Frackowiak, R. S., and Frith, C. D. 1995. Other minds in the brain: A functional imaging study of "theory of mind" in story comprehension. *Cognition* **57:** 109–128.

Flynn, F., Cummings, J. L., and Tomiyasu, U. 1988. Altered behavior associated with damage to the ventromedial hypothalamus: A distinctive syndrome. *Behavioral Neurol.* **1:** 49–58.

Forman, S. D., Cohen, J. D., Fitzgerald, M., Eddy, W. F., Mintun, M. A., and Noll, D. C. 1995. Improved assessment of significant activation in functional magnetic resonance imaging (fMRI): Use of a cluster-sized threshold. *Magn. Reson. Med.* **33:** 636–647.

Friston, K. J., Holmes, A. P., Poline, J.-P., Grasby, P. J., Williams, S. C., Frackowiak, R. S., and Turner, R. 1995. Analysis of fMRI time-series revisited. *Neuroimage* **2:** 45–53.

Frith, C. D., and Frith, U. 1999. Interacting minds—A biological basis. *Science* **286:** 1692–1695.

Gilligan, C. 1993. *In a Different Voice,* 3rd ed. Harvard, Cambridge.

Golby, A. J., Gabrieli, J. D. L., Chiao, J. Y., and Eberhardt, J. L. Differential responses in the fusiform region to same-race and other-race faces. *Nature Neurosci.* **4:** 845–850.

Gorman, D. G., and Cummings, J. L. 1992. Hypersexuality following septal injury. *Arch. Neurol.* **49:** 308–310.

Grafman, J., Schwab, K., Warden, D., Pridgen, A., Brown, H. R., and Salazar, A. M. 1996. Frontal lobe injuries, violence, and aggression: A report of the Vietnam Head Injury Study. *Neurology* **46:** 1231–1738.

Greene, J. D., Sommerville, R. B., Nystrom, L. E., Darley, J. M., and Cohen, J. D. 2001. An fMRI investigation of emotional engagement in moral judgment. *Science* **293:** 2105–2108.

Haidt, J. The moral emotions. In *Handbook of Affective Sciences* (R. J. Davidson, K. Scherer, H. H. Goldsmith, Eds.) Oxford Univ. Press.

Hoffman, E. A., and Haxby, J. V. 2000. Distinct representations of eye gaze and identity in the distributed human neural system for face perception. *Nature Neurosci.* **3:** 80–84.

Isenberg, N., Silbersweig, D., Engelien, A., Emmerich, S., Malavade, K., Beattie, B., Leon, A. C., and Stern, E. 1999. Linguistic threat activates the human amygdala. *Proc. Natl. Acad. Sci. USA* **96:** 10456–10459.

Kandel, E., and Freed, D. 1989. Frontal-lobe dysfunction and antisocial behavior: a review. *J. Clin. Psychol.* **45:** 404–413.

Kiehl, K. A., Smith, A. M., Hare, R. D., Mendrek, A., Forster, B. B., Brink, J., and Liddle, P. F. 2001. Limbic abnormalities in affective processing by criminal psychopaths as revealed by functional magnetic resonance imaging. *Biol. Psychiatry* **50:** 677–684.

Lane, R., Chua, P., and Dolan, R. 1999. Common effects of emotional valence, arousal and attention on neural activation during visual processing of pictures. *Neuropsychologia* **37:** 989–997.

Markowitsch, H. J. 1998. Differential contribution of right and left amygdala to affective information processing. *Behav. Neurol.* **11:** 233–244.

Miller, B. L., Hou, C., Goldberg, M., and Mena, I. 1999. Anterior temporal lobes: Social brain. In *The Human Frontal Lobes: Functions and Disorders.* (B. L. Miller and J. L. Cummings, Eds.) Guilford Press.

Moll, J., Eslinger, P. J., and 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

Morris, J. S., Friston, K. J., Buchel, C., Frith, C. D., Young A. W., Calder A. J., and Dolan R. J. 1998. A neuromodulatory role for the human amygdala in processing emotional facial expressions. *Brain* **121:** 47–57.

O'Doherty, J., Kringelbach, M. L., Rolls, E. T., Hornak, J., and Andrews, C. 2001. Abstract reward and punishment representations in the human orbitofrontal cortex. *Nature Neurosci.* **4:** 95–102.

Oliveira-Souza, R., and Moll, J. 2000. The moral brain: Functional MRI correlates of moral judgment in normal adults. *Neurology* **54** (Suppl. 3): 252.

Öngür, D., and Price, J. L. 2000. The organization of networks within the orbital and medial prefrontal cortex of rats, monkeys and humans. *Cereb. Cortex* **10:** 206–219.

Phelps, E. A., O'Connor, K. J., Gatenby, J. C., Gore, J. C., Grillon, C., and Davis, M. 2001. Activation of the left amygdala to a cognitive representation of fear. *Nature Neurosci.* **4:** 437–441.

Pietrini, P., Guazzelli, M., Basso, G., Jaffe, K., and Grafman, J. 2000. Neural correlates of imaginal aggressive behavior assessed by positron emission tomography in healthy subjects. *Am. J. Psychiatry* **157:** 1772–1781.

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

Reiman, E. M., Lane, R. D., Ahern, G. L., Schwartz, G. E., Davidson, R. J., Friston, K. J., Yun, L. S., and Chen, K. 1997. Neuroanatomical correlates of externally and internally generated human emotion. *Am. J. Psychiatry* **154:** 918–925.

Rolls, E. T., Hornak, J., Wade, D., and 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.

Ruby, P., and Decety, J. 2001. Effect of subjective perspective taking during simulation of action: A PET investigation of agency. *Nature Neurosci.* **4:** 546–550.

Snarey, J. R. 1985. Cross-cultural universality of social-moral development: A critical review of Kohlbergian research. *Psychol. Bull.* **97:** 202–232

Talairach, J., and Tournoux, P. 1998. *Co-Planar Stereotaxic Atlas of the Human Brain.* Thieme Medical, New York.

Zahn, R., Huber, W., Drews, E., Erberich, S., Krings, T., Willmes, K., and Schwarz, M. 2000. Hemispheric lateralization at different levels of human auditory word processing: A functional magnetic resonance imaging study. *Neurosci. Lett.* **287:** 195–198