The Neural Basis of Mood-Congruent Processing Biases in Depression

Rebecca Elliott, PhD; Judy S. Rubinsztein, MB ChB, MRCPsych; Barbara J. Sahakian, PhD, DipClinPsych; Raymond J. Dolan, MD, PhD

Background: Mood-congruent processing biases are among the most robust research findings in neuropsychological studies of depression. Depressed patients show preferential processing of negatively toned stimuli across a range of cognitive tasks. The present study aimed to determine whether these behavioral abnormalities are associated with specific neural substrates.

Methods: Ten depressed patients and 11 healthy control subjects underwent scanning during performance of an emotional go/no-go task using functional magnetic resonance imaging. The task allowed comparison among neural response to happy, sad, and neutral words, in the context of these words as targets (ie, stimuli to which subjects were required to make a motor response) or distractors (ie, stimuli to which the motor response was withheld).

Results: Depressed patients showed attenuated neural responses to emotional relative to neutral targets in ventral cingulate and posterior orbitofrontal cortices. However, patients showed elevated responses specific to sad targets in rostral anterior cingulate extending to anterior medial prefrontal cortex. Unlike controls, patients showed differential neural response to emotional, particularly sad, distractors in the lateral orbitofrontal cortex.

Conclusions: These findings suggest a distinct neural substrate for mood-congruent processing biases in performance. The medial and orbital prefrontal regions may play a key role in mediating the interaction between mood and cognition in affective disorder.

Arch Gen Psychiatry. 2002;59:597-604

A RELIABLE FINDING in neuropsychological studies of depression is a bias toward processing of mood-congruent information. Depressed patients show a facilitation of performance when responding to stimuli with a negative emotional tone. This phenomenon has been observed in various cognitive contexts. Studies of memory in depressed patients have reported a tendency toward recall of negatively toned material,1-5 and have argued that these memory biases may be, at least partly, unconscious and therefore evident when memory is studied implicitly.6 Mood-congruent biases have also been demonstrated in attentional paradigms.7,8 For example, depression-related words cause significantly greater interference in Stroop tasks than neutral or happy words.9,10 A recent study of an emotional go/no-go task also found a bias toward sad stimuli in depressed patients11 and demonstrated a contrasting bias toward positive information in patients with manic-depressive disorder.

Although mood-congruent biases have been reliably demonstrated, their exact relationship to depressive symptomatology is unclear. The approach of studying emotional biases has the advantage of explicitly linking mood and cognition in a manner that can be related to cognitive-behavioral theories of depression, on which treatment strategies have been based.11 Thus, patients with depressed mood may be differentially sensitive to negatively toned information and process it more effectively. This finding could reinforce depressed mood and contribute to the maintenance of the disorder. A key question in understanding the role of mood-congruent processing biases in depression concerns the neural basis for this phenomenon. A previous study of an emotional Stroop paradigm12 suggested that functional abnormalities of the anterior cingulate cortex are involved in mood-congruent response biases in depression, and we sought to examine the issue further.

In a recent study,13 we used a version of the emotional go/no-go task developed by Murphy et al10 in a functional magnetic resonance imaging (MRI) study of control subjects to assess differential neural responses to the emotional valence of verbal
SUBJECTS AND METHODS

SUBJECTS

Ten patients with a diagnosis of unipolar recurrent major depression were recruited from an affective disorders clinic at Addenbrookes Hospital, Cambridge, England (7 women and 3 men). Nine patients were right-handed and 1 was left-handed. The diagnosis was established using a structured interview (the Schedule for Affective Disorders and Schizophrenia–Lifetime Version) and case note review. Patients had to fulfill research diagnostic criteria and DSM-IV criteria for major depressive disorder at each episode, and those with a history of neurologic disease or closed head injury were excluded. Diagnosis was assigned by one of us (J.S.R.). No patients with current comorbid anxiety disorders, substance abuse or dependence, or other psychiatric diagnoses based on DSM-IV were included. None of the patients had bipolar disorder. One patient had a history of panic disorder and another had a history of bulimia, but neither fulfilled DSM-IV criteria for these disorders at the time of the study. Patients were also excluded if they were not euthyroid or if they had histories of other endocrine disorders or unstable medical disorders. Two patients were receiving treatment for asthma, 1 for Behçet disease, and 1 for a hiatal hernia, but these conditions were stable at the time of scanning. One female patient was postmenopausal and was receiving hormone replacement therapy; no patient was receiving hormonal contraceptives.

Patients were aged 30 to 59 years, with a mean age of 42.2 years (SD, 8.3 years). Severity of depression was assessed using the 17-item Hamilton Depression Scale (mean score, 23.1; SD, 3.9; range, 17-30) and the Montgomery-Asberg Depression Rating Scale (mean score, 31.3; SD, 5.2; range, 28-42). Only 1 patient was currently hospitalized. The mean time since first diagnosis was 15 years (SD, 3.3 years; range, 8-23 years), and the mean number of depressive episodes was 3.2 (SD, 1.2; range, 2-5). All patients were receiving medication, and had been for 1 to 8 years. Four patients were receiving tricyclic antidepressants; 4, selective serotonin reuptake inhibitors; 1, venlafaxine hydrochloride; and 1, nefazadone hydrochloride. In addition, 3 patients were receiving lithium carbonate; 2, antipsychotics (100 mg of oral thioridazine hydrochloride daily); and 3, long-term benzodiazepine therapy. The clinical heterogeneity observed in the syndrome of major depression was deliberately limited by including only patients with recurrent major depression and those under the care of a psychiatrist within the secondary health care system in the United Kingdom. In all cases, despite medication, patients were clinically depressed.

These patients were compared with 11 right-handed volunteers who were recruited by means of advertisement in the local community and underwent MRI scanning in our earlier study (8 women and 3 men; aged 24-59 years [mean age, 37.6 years; SD, 9.7 years]). No significant difference was found in age between control subjects and patients. Controls underwent screening using a verbal interview, the Beck Depression Inventory, and the General Health Questionnaire to exclude any current depressive symptomatology, neurologic or psychiatric history, closed head injury, or substance abuse. One female control was postmenopausal and was receiving hormone replacement therapy. Our results were not affected by removing this control and the corresponding patient from the analysis. Three of the controls were receiving oral contraceptives.

The study was approved by local research ethics committees (Joint Ethics Committee of National Hospitals and Institute of Neurology, London, England, and the Addenbrookes Hospital Research Ethics Committee). Informed written consent was obtained from all subjects.

COGNITIVE ACTIVATION PARADIGM

The cognitive activation paradigm is discussed in detail in Elliott et al and was identical for both subject groups. Twenty-four task blocks were interspersed with 24 rest blocks. In each block, subjects performed a variant of a classic go/no-go task. Before the start of a block, subjects were given an instruction to respond to certain targets (go) but ignore distractors (no-go). In the main task conditions, subjects responded to happy, sad, or neutral targets. The words in each category were matched for imageability, word length, and frequency. Representative examples included joyful,
success, and confident for happy; gloomy, hopeless, and failure for sad; and range, vary, and directly for neutral. Under control conditions, all of the words were neutral, and the targets were defined on the basis of font (italic vs plain text, an orthographic control condition). Overall, we used the following 8 conditions: (1) happy targets and sad distractors; (2) happy targets and neutral distractors; (3) sad targets and happy distractors; (4) sad targets and neutral distractors; (5) neutral targets and happy distractors; (6) neutral targets and sad distractors; (7) all words neutral, with targets in italic and distractors in plain text; and (8) all words neutral, with targets in plain text and distractors in italic.

In conditions 1 through 6, the task involved judging whether word stimuli were of one or the other emotional valence. For instance, in condition 1, subjects were instructed to respond with a button press to happy words (targets), and to withhold the press for sad words (distractors). Thus, conditions 1 through 6 assessed the effects of attending to words of different emotional tone and allowed differential responses to both target and distractor valence to be assessed. Conditions 7 and 8 were control conditions in which subjects were not required to make semantic judgments.

In each block, 10 targets and 10 distractors were presented in a randomized order. Each word appeared for 300 millisecondst, and an interstimulus interval of 900 milliseconds allowed subjects to respond (or not). Each 20-word block was 24 seconds long and was preceded by a 24-second rest block. At 4 seconds before the end of each rest block, a written instruction for the next task block appeared on the screen. Subjects responded by pressing a button as quickly as possible every time they detected a target. Equal numbers of happy and sad words were seen during scanning, and therefore we did not anticipate any effect of mood induction on task performance. However, we examined whether a significant time × condition interaction existed, which would reflect a systematic mood change.

**MRI SCANNING**

We acquired MRI data using a 2-T system (Siemens VISION; Siemens AG, London). We acquired functional images by means of a gradient echo, echo-planar T2* sequence using blood oxygenation level dependency contrast. We obtained 294 functional images for each subject. Each image constituted a full brain volume of 48 axial slices at 3-mm separation and with 3 mm in plane resolution, acquired continuously with a repetition time of 4 seconds. The first 6 volumes were dummy volumes to allow for T1-weighted equilibration effects, followed by 6 volumes per block. We also acquired T1-weighted structural images for each subject.

**DATA ANALYSIS**

Data were analyzed using statistical parametric mapping (SPM99; Wellcome Department of Cognitive Neurology, London, England), which was implemented using MATLAB (The Mathworks Inc, Sherborn, Mass), and run on a SPARC workstation (Sun Microsystems, Inc, Surrey, England). This approach to the analysis of functional imaging data has been described in detail elsewhere.\(^{20,21}\) In brief, scans were first realigned, normalized, and spatially smoothed to correct for subject motion and to facilitate intersubject averaging. A random-effects statistical model was used to analyze the data, which accounted for intrasubject variability and allowed population-based inferences to be drawn. For each subject, 1 mean image per condition was generated, and these were combined in a series of linear contrasts to assess group effects. These comparisons generated statistical parametric maps (SPMs) of the \(t\) statistic (SPM\(t\)), which was transformed to a normal distribution (SPM\(Z\)). In line with established functional imaging conventions, we report neural responses seen at an uncorrected threshold of \(P<.001\) for regions about which we had an a priori hypothesis. These regions were the ventral and medial prefrontal regions, limbic system, thalamus, and insula. For descriptive purposes, we also report neural responses at this threshold in regions for which there was no prior hypothesis. However, we restrict discussion and interpretation of the regions with no hypothesis to those that survived the more stringent threshold of \(P<.05\), corrected for multiple comparisons. The designation of anatomical localizations are based on the structural MRIs of the group and the atlas of Duvernoy.\(^{24}\)

We compared conditions 1 and 3 with conditions 5 and 6. Conditions 1 and 3 match conditions 5 and 6 for distractor valence, but differ in terms of target valence (happy and sad vs neutral). In several regions, we found a significant group × condition interaction (Table 2 and Figure 1) for the comparison of emotional relative to neutral words in patients relative to controls. These regions included the left inferior frontal gyrus, ventral cingulate cortex (extremely close to the subgenual region), left middle temporal gyrus, left precentral gyrus, left postcentral gyrus, and bilateral pulvinar region of the thalamus. As shown in Figure 1D, neural responses to neutral words in these regions was comparable for both groups, but an enhanced response under the emotional conditions, observed in controls, was not seen in patients. The interaction is thus driven by attenuation of an emotion-specific response in depressed patients. Figure 1B also shows that the emotion-specific response in controls was more pronounced for happy than for sad words.

A significant differential group response was also seen in the reverse contrast, involving an extensive region of the right lateral frontal cortex (Brodmann area [BA] 9/46; Table 2 and Figure 2). As shown in Figure 2B, this group × condition interaction was driven by an elevated response to neutral words in controls and by an elevated response to sad words in depressed patients.

**HAPPY COMPARED WITH SAD TARGETS**

We compared conditions 2 and 4 (matched for distractor valence and differing only in terms of target valence).
A significant interaction between emotional valence and subject group was seen in a region of medial prefrontal cortex extending from the rostral anterior cingulate (BA 32/24) anteriorly to the medial prefrontal cortex (BA 9/10) (Table 2 and Figure 3). This interaction was driven by a relatively enhanced response to sad targets in patients and to happy targets in controls. We also found a significant response in the right anterior temporal lobe (BA 38), left middle temporal gyrus (BA 21), and bilateral medial frontal gyrus (BA 6). Again, these interactions were driven by a relatively enhanced response to happy targets in controls and to sad targets in patients.

No regions were found where controls showed a greater response to sad targets or depressed patients showed a greater response to happy ones.

**EFFECTS OF DISTRACTORS**

Relative to controls, depressed patients showed an enhanced neural response to emotional compared with neutral distractors (Table 3). This contrast represents conditions 1 and 3 compared with conditions 2 and 4 (matched target valence and different distractor valence). Regions showing this enhanced response were the bilateral lateral orbitofrontal cortices (OFC) (BA 11/47) and bilateral anterior temporal lobe (BAs 20 and 38).

When the different distractor valences were compared (using conditions 5 and 6 to match for target valence), we found a greater response to sad than to happy distractors in the right lateral OFC (BA 11/47) and the bi-
lateral anterior temporal lobe (Table 3 and Figure 4) in patients but not in controls. Enhanced responses in the same regions were seen when sad distractors were compared with neutral distractors with matched happy targets.

We found no significant time × condition interactions, suggesting that neural responses under the different conditions were consistent across time. This finding fulfills the expectation that no systematic effect of induced mood and no significant learning effect on this task existed.

CORRELATIONS WITH DEPRESSION SEVERITY

We found no regions where condition-specific neural responses in depressed patients correlated significantly with depression severity.

The key findings of this study are abnormal neural responses associated with emotional processing biases in depressed patients in regions including the medial and
ventral prefrontal cortices. Depressed patients showed a general attenuation of neural responses to emotional words in cortical and subcortical structures, including a ventral cingulate region adjacent to the subgenual cingulate. More specifically, we found a double dissociation of valence-specific response in the rostral anterior cingulate cortex extending to the anterior medial prefrontal cortex. This region responded more strongly to happy words in controls and, conversely, to sad words in patients. Finally, we found differential responses to the emotional valence of distractor words in patients, but not in controls. A right lateral orbitofrontal response was associated with the presence of sad distractors.

Attenuated neural responses to emotional targets in depressed patients were seen in several regions, including a ventral region of anterior cingulate cortex, close to the subgenual region discussed in seminal studies by Drevets et al.14,15 Attenuation was more pronounced for happy than for sad targets. This region is a focus of structural and functional abnormalities in patients with depression, and the present finding suggests that a functional abnormality here might mediate emotional modulation of cognitive processing. A similar pattern of attenuated response to emotional targets was seen in posterior OFC, insula, and thalamus. The OFC25 and the insula26,27 have been implicated in representing changes in body state associated with emotional responses, and in monitoring autonomic responses to emotionally salient information.28,29 The thalamus has been extensively linked to attention and arousal mechanisms.30 Attenuated responses in these interconnected regions of ventral cingulate, posterior OFC, insula, and thalamus may therefore suggest a failure of normal autonomic emotional arousal mechanisms in depressed patients.

We also found a region of right dorsolateral prefrontal cortex (DLPFC) where depressed patients showed greater response to emotional targets than did controls. This finding was due to depressed subjects showing enhanced response to sad words, whereas controls showed enhanced response to neutral words. The DLPFC is not considered a classic substrate for emotional aspects of processing. Rather, it has been implicated in a number of higher cognitive functions, including working memory,31,32 episodic memory retrieval,33 attentional set shifting,34,35 planning,36 and monitoring.37,38 Of these processes, monitoring is most obviously involved in the present task, with subjects required to monitor a stream of input for targets. In controls, the monitoring processes subserved by right DLPFC may be biased toward neutral targets; however, for patients, these monitoring processes may be biased toward mood-congruent (sad) targets. George et al11 also reported enhanced DLPFC response in depressed patients performing an emotional Stroop task with sad words, although this was left lateralized.

The classic finding of the response bias in the literature is not only a bias toward sad information in depression, but also a bias toward happy information in controls.3 This behavioral dissociation has direct correlates in the neural response of a medial prefrontal region, extending from the rostral cingulate to the anterior medial prefrontal cortex, and also the right anterior temporal lobe. These regions respond differentially to happy targets in controls but to sad targets in depressed patients. A very similar prefrontal region, spanning the rostral cingulate and anterior medial prefrontal cortices, was activated in a positron emission tomographic study of subjective emotional responses.39 The present finding suggests that this subjective emotional system is biased toward mood-congruent information in both controls and depressed patients. Like the subgenual cingulate cortex, this more rostral and anterior region has been critically associated with the neuropathology of depression. Mayberg et al40 described the overall level of metabolism in this region as a potential predictor of treatment response. The present results suggest that a functionally abnormal response in this region may mediate the bias of depressed patients toward negative information that is considered by cognitive theorists to be crucial in the maintenance of the disorder.11,41

A final important finding of this study is that, unlike controls, depressed patients make differential responses to emotional distractors in bilateral anterior temporal regions and lateral orbitofrontal regions. In a direct comparison between sad and happy distractors, we found an enhanced neural response to sad distractors in the right lateral OFC. This finding suggests that irrelevant sad material has a greater capacity to capture attention in depressed patients. We found no performance differences, suggesting that the effect operates at a neural but not a behavioral level. Right lateral OFC response has been associated with performance of nonemotional go/no-go tasks.42,43 Arguments have been made44 that this region critically mediates behavioral inhibition mechanisms and is therefore recruited when subjects are required to withhold a prepared or prepotent response. The present findings suggest that in depressed subjects, mood-congruent distractors place greater demands on these inhibitory processes in the emotional go/no-go task.

The differential responses to particular emotional conditions observed herein occurred in the absence of significant performance differences or general task-related attenuation of neural responses. Thus, this finding obviates interpretational problems that typically plague neuroimaging studies of psychiatric disorders. Condition-specific differences in neural response cannot be attributed to confounds arising from significant performance impairment, since depressed patients perform as accurately as controls. Also, when all of the active conditions in the task are compared with rest conditions, or indeed when semantic conditions are compared with orthographic conditions, no significant differences are found in neural response for patients relative to controls. The condition-specific differences, therefore, do not have to be interpreted in the context of generalized attenuation in response to cognitive challenge.

However, a number of limitations apply to the present study. One potential confound is possible drug effects. All patients taking part in the study were receiving medication, but the variety of different medication regimens precluded any attempt to compare the effects of different drugs. However, there was an absence of group differences when all the active conditions were compared with the rest or the lower-level control condition, which is not consistent with any systematic effect of the drug on cognitive activation. Nevertheless, all subjects were medication nonresponders and therefore repre-
sent a specific subgroup of patients with depression. Future study with different subgroups, including medication-free patients, is needed to establish the generalizability of the effects observed here. The findings reported herein may be generalized to other tasks assessing processing biases. For example, Whalen and colleagues have developed an emotional version of the Stroop task for functional MRI and, as with our go/no-go task, have reported a crucial role for the anterior cingulate in healthy subjects. George et al. have reported anterior cingulate cortex abnormalities in depressed patients performing the emotional Stroop task during a single-photon emission computed tomography study. The clear prediction is that this abnormal anterior cingulate cortex response to emotional words would be seen in functional MRI findings, reflecting mood-congruent biases. An event-related approach would also disambiguate the abnormal responses to specific targets and distractors in response bias paradigms, which are inevitably confounded to some extent in the blocked approach used in this study.

To our knowledge, this study is one of the first to look at the functional neuroanatomy of the emotional modulation of cognitive processing in depressed patients. Previous studies have reported abnormalities in the network subserving passive viewing of emotional material that responds to therapeutic intervention. However, specific studies of the interface between mood and cognition have potentially important implications for our understanding of affective disorders. Distortions of cognitive processing by affective factors are at the core of many influential theories of depression, and functional neuroimaging studies provide the means to relate these distortions to neurobiological abnormalities. The present findings suggest a critical involvement of ventral and medial prefrontal regions in mediating mood-congruent response biases in depression. We suggest that distinct regions mediate separable aspects of this effect and hypothesize a testable theoretical framework.

Submitted for publication September 21, 2000; final revision received May 30, 2001; accepted August 13, 2001.

This study was supported by a program grant to Dr Dolan and to T. W. Robbins, PhD, B. J. Everitt, PhD, A. C. Roberts, PhD, and Dr Sahakian from Wellcome Trust, London; by the Betty Behrens research fellowship at Clare Hall, Cambridge (Dr Rubinsztein); and by a Sackler studentship, University of Cambridge Medical School, Cambridge (Dr Rubinsztein).

Corresponding author and reprints: Rebecca Elliott, PhD, Neuroscience and Psychiatry Unit, Room G907, St. John's College, University of Cambridge, Cambridge, CB2 1ST, England (e-mail: rebecca.elliott@man.ac.uk).

REFERENCES

33. Rugg MD, Fletcher PC, Frith CD, Frackowiak RSJ, Dolan RJ. Differential activa-


