Psychiatry Research, 42:241-251
Elsevier

Facial Emotion Discrimination: II. Behavioral Findings in Depression

Ruben C. Gur, Roland J. Erwin, Raquel E. Gur, Alexander S. Zwil, Carolyn Heimberg, and Helena C. Kraemer

Received September 17, 1991; revised version received February 24, 1992; accepted April 12, 1992.

Abstract. The facial discrimination tasks described in part I (Erwin et al., 1992) were administered to a sample of 14 patients with depression and 14 normal controls matched for sex (12 women, 2 men) and balanced for age and sociodemographic characteristics. Patients performed more poorly on measures of sensitivity for happy discrimination and specificity for sad discrimination, and had a higher negative bias across tasks. Severity of negative affect was correlated with poorer performance for patients. The results suggest that depression is associated with an impaired ability to recognize facial displays of emotion.

Key Words. Affective disorder, anxiety, mood, human performance.

Emotional dysfunction is an important feature of psychopathology. However, while abnormalities in the expression of emotion are used diagnostically, the ability to recognize and discriminate affective states in others has not been well characterized. Impairment in this skill could be linked to poor social and interpersonal relations, a feature of psychopathology.

While affective disturbances are central to depression, studies that have examined facial affect recognition in patients with psychopathology have focused on schizophrenia (for review, see Morrison et al., 1988). Patients with affective illness have often been used as a psychiatric control group (Cutting, 1981; Feinberg et al., 1986; Zuroff and Colussy, 1986; Gessler et al., 1989). Some studies have reported that both patients with depression and patients with schizophrenia perform more poorly than normal controls, while others have reported that schizophrenia, but not depression, is associated with poor performance. The inconsistencies are partially due to variation in subject characteristics (see Morrison et al., 1988), but a number of methodological issues related to the stimuli and procedures also play a role (Erwin et al., 1992).

If emotional discrimination deficits were a component of disturbed affect in depression, samples of patients closely matched to controls would reveal difficulties

Ruben C. Gur, Ph.D., and Raquel E. Gur, M.D., Ph.D., are Professors; Roland J. Erwin, Ph.D., is Assistant Professor; and Alexander S. Zwil, M.D., and Carolyn Heimberg, M.D., are Postdoctoral Fellows, all in the Department of Psychiatry, University of Pennsylvania, Philadelphia, PA, at the time this report was submitted. Helena Kraemer, Ph.D., is Professor, Department of Psychiatry and Behavioral Sciences, Stanford University, Stanford, CA. Alexander Zwil, M.D., is now at the Department of Psychiatry, Thomas Jefferson University, Philadelphia, PA. (Reprint requests to Dr. R.C. Gur, Brain-Behavior Laboratory, Dept. of Psychiatry, 10th Floor Gates Bldg., University of Pennsylvania, 3500 Spruce St., Philadelphia, PA 19104-4283, USA.)

either in the form of performance deficits (i.e., sensitivity or specificity) or a criterion shift (bias). If such deficits existed and were germane to the clinical symptoms, they would relate to the severity of depression. In part I (Erwin et al., 1992), we described the development of an emotional discrimination task and its application in normal subjects. The task seemed sensitive to sex differences in emotional discrimination and hence could be used to examine possible disruption in facial discrimination associated with psychopathology. Here we present performance data in patients with depression and normal controls matched on age, sex, and ethnic background. We test the following hypotheses: (1) Patients are impaired in emotional discrimination. (2) Patients have a negative bias to judge neutral faces as sad or happy faces as neutral. (3) The degree of abnormality in performance of patients is correlated with the severity of negative affect.

Methods

Subjects. Patients with depression (n = 14; mean age = 44.5 years, SD = 12.9, range = 24-71) and normal controls (n = 14; mean age = 36.6 years, SD = 16.32, range = 20-73) participated. Patients were recruited from the inpatient (n = 9) and outpatient (n = 5) facilities of the Affective Disorders Program of the University Hospital and met *DSM-III-R* (American Psychiatric Association, 1987) criteria for major depression or bipolar illness, depressed phase. Table 1 presents patient characteristics.

The average age of onset for the bipolar patients was 18.0 years, and that for unipolar depressive patients was 37.3 years. Bipolar patients had numerous episodes of illness as they included three rapid cyclers and had an average of 1.8 prior hospitalizations. Unipolar

Table 1. Patient characteristics

Patient	Age/Sex	Diagnosis	HRSD	Medication
1	59 F	B-I,D	28	Nortriptyline, carbamazepine, thyroxine
2	36 F	B-I,D	34	Lithium
3	24 F	B-II,D	31	Diazepam, trifluoperazine
4	44 M	B-II,D	28	Lithium, diazepam, fluoxetine, methylphenidate
5	50 F	B-II,D	19	Lithium
6	40 F	D	22	None
7	46 F	D	22	Lithium, clonazepam, phenelzine
8	38 F	D	18	None
9	53 M	D	23	Desipramine (1 dose)
10	31 F	D,PSY	23	Nortriptyline
11	44 F	D,PSY	26	Nortriptyline, trifluoperazine
12	29 F	D,PSY	22	Trazodone, thiothixene, benztropine
13	71 F	D,PSY	28	Nortriptyline, thiothixene, alprazolam, diphenhydramine
14	57 F	D,NOS	29	Desipramine, lorazepam

Note. B-I,D = bipolar I, depressed. B-II,D = bipolar II, depressed. D = major depression. D,PSY = major depression with psychotic features. D,NOS = depressive disorder not otherwise specified. HRSD = Hamilton Rating Scale for Depression. F = female. M = male.

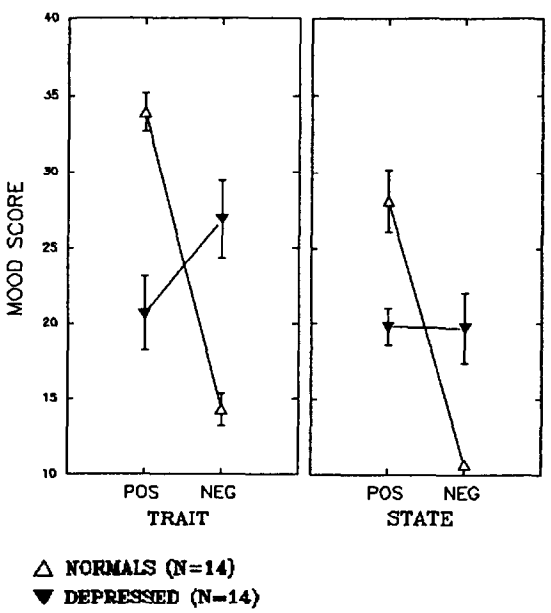
patients had 2.6 previous episodes and 1.6 prior hospitalizations. Among bipolar patients, all had a positive family history of psychiatric illness, four with mood disorder and four with alcoholism. Among unipolar patients, one had a family history of mood disorder, and three of alcoholism.

The control subjects are a subsample of the data reported in part I (Erwin et al., 1992). Subjects were selected to balance patients for age and were matched for sex (12 women and 2 men per group), education, and ethnicity. No significant difference in number of years of education was found across groups (controls: mean = 15.5 years, SD = 2.6; patients: mean = 14.3 years, SD = 2.6; $t = 1.24$, NS).

Clinical and Mood Scales. Subjects were administered trait and state mood (Positive and Negative Affect Scale [PANAS] pre, post, and general; Zevon and Tellegen, 1982) and anxiety (State-Trait Anxiety Inventory [STAI]; Spielberger et al., 1970; state pre and post and trait) scales, and a self-report depression scale (Beck Depression Inventory [BDI]; Beck et al., 1961). The Hamilton Rating Scale for Depression (HRSD; Hamilton, 1960) was also used to assess the depressed patients (mean score = 25.5, SD = 4.7, range = 18-35). Patients had higher BDI scores than controls (patients: mean score = 26.4, SD = 8.9; controls: mean score = 4.1, SD = 1.1; $t = 8.32$, $df = 16.4$ [unequal variances], $p < 0.001$).

Analysis of PANAS trait and state scores showed a diagnosis × positive-negative affect interaction for both trait ($F = 29.22$; $df = 1, 25$; $p < 0.0001$) and state ($F = 23.03$; $df = 1, 25$; $p < 0.0001$). Patients had equal scores for positive and negative affect, while normal subjects had higher scores for positive affect and lower scores for negative affect (Fig. 1) for both trait (1a) and state (1b).

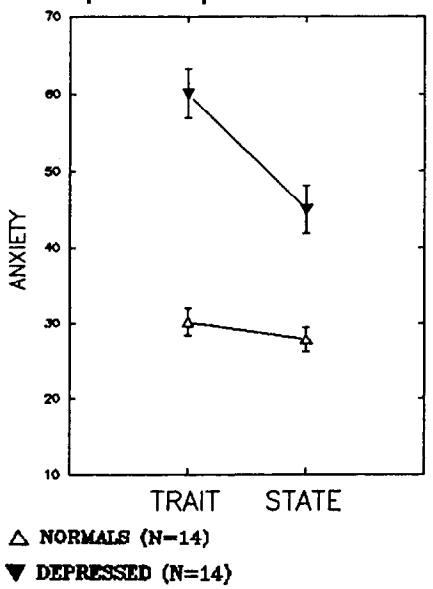
Fig. 1. PANAS positive and negative mood scores (means and standard errors) of patients and controls for both trait (PANAS GENERAL) and state (PANAS NOW) scales



PANAS = Positive and Negative Affect Scale.

The STAI showed a main effect for diagnosis ($F = 92.47$; $df = 1, 25$; $p < 0.0001$), with patients having overall higher anxiety. The STAI also showed a trait-state × diagnosis interaction ($F = 6.21$; $df = 1, 25$; $p < 0.025$), with normal subjects having equal trait and state scores, and patients scoring higher on trait than state measures (Fig. 2).

Fig. 2. State-Trait Anxiety Inventory scores (means and standard errors) of normal controls and depressive patients



Task Administration. The tasks were described in part I (Erwin et al., 1992). In brief, the stimuli consisted of faces expressing degrees of happy and sad emotion as well as neutral expressions. There were three discrimination tasks: age, happy-neutral, and sad-neutral. During the age-discrimination task, subjects were instructed to indicate the age of the poser by decade (1 = teens, 2 = twenties, etc.). For the happy-neutral and sad-neutral tasks, subjects were instructed to indicate on a 1- to 7-point scale whether the face presented was very happy, happy, mildly happy, neutral, mildly sad, sad, or very sad. The polarity of the scale for the emotion-discrimination tasks was counterbalanced across tasks and subjects.

Stimuli (35-mm slides) were projected on a white background, and subjects indicated their answers with a flashlight pointer held in both hands. Stimulus duration was 7 seconds. Each task was self-paced over a 10-minute interval, and responses were obtained for each stimulus before the next trial was initiated. Subjects were presented same-gender stimulus sets (i.e., males saw male faces).

Data Analysis. True positives (i.e., response in the happy range for a happy slide in the happy discrimination task and in the sad range for a sad slide in the sad discrimination task), false positives (scores in the emotional range for neutral stimuli), true negatives (neutral responses to neutral faces), and false negatives (neutral responses to emotional faces) were counted for each task as described in part I (Erwin et al., 1992). Percent correct, sensitivity, and specificity measures were also computed (Erwin et al., 1992). In addition, a positive bias across tasks was computed as (happy-neutral false positives + sad-neutral false negatives)/(total attempted happy-neutral positives + total attempted sad-neutral negatives), and negative bias was computed as (happy-neutral false negatives + sad-neutral false positives)/(total attempted happy-neutral negatives + total attempted sad-neutral positives).

The formula 2*arc sine of the square root of the proportion, as recommended by Cohen (1988), was used to transform all proportions. To remove the effects of age, sex, and the age + sex interaction, the following procedures were used: (1) β weights were generated based on regression of these variables (age, sex, and age*sex) on the performance scores of the entire normative sample (24 men and 15 women) included in part I's normative analyses (Erwin et al., 1992). (2) For each performance score, the β weight of the corresponding adjusting variable (e.g., age) was multiplied by each subject's value for that adjusting variable standardized to a mean of zero (based on the normative sample). (3) The resulting values for each adjusting variable were then subtracted from the original scores (for a further discussion of this technique, see Saykin et al., submitted). Thus, the performance data submitted for statistical analysis were not influenced by effects of age and sex on normal performance. These regression adjustments were preferred over using age and sex as covariates in an omnibus analysis of covariance (ANCOVA) since they allow information on the total sample of normal subjects to be used in accounting for sex and age effects. This provides more stable estimates, while an ANCOVA would use only subjects included in the current analysis. Also, the regression adjustments are not influenced by possible diagnosis $\times$ covariate interactions, as is the case with ANCOVA.

Statistical analysis was performed in two stages. In the first stage, the specific hypotheses were evaluated. To evaluate hypothesis 1 (i.e., patients are impaired in emotional discrimination), mixed model repeated measures analyses of variance (ANOVAs; SAS Institute, 1985), with diagnosis as a grouping factor and task (happy-neutral, sad-neutral) as the within-group (repeated measures) factor, were used to compare sensitivity and specificity scores between patients and controls. To evaluate hypothesis 2 (i.e., patients have a negative bias to judge neutral faces as sad or happy faces as neutral), the negative and positive bias scores were entered as dependent measures in this ANOVA design, and a planned contrast was performed for the negative bias score. Hypothesis 3 (i.e., the degree of abnormality in performance of patients is correlated with the severity of negative affect) could be tested by correlating the performance and bias measures with the PANAS, BDI, and HRSD scores. In view of the sample size and the power of product-moment correlations, however, calculating all these would have inflated the chance of type I errors. We therefore correlated the performance measures and the negative bias measure only with the PANAS Negative Affect scale (trait measure) for the entire sample and the HRSD for patients, and used $p < 0.01$ as a cutoff for significance.

In the second stage of the statistical analysis, hypotheses 1 and 2 were contrasted to address the question of which derived measure best discriminated patients from controls. Kraemer's Quality Receiver Operating Characteristic (QROC) method (Kraemer, 1988) was applied to the sensitivity, specificity, and bias scores. Also, traditional analyses were conducted on the percent correct scores for all tasks, and the mean mood ratings for the first and second half of responses during each task (the halfway point was determined by dividing the subject's total number of ratings on each task by 2). These mood rating data were analyzed by adding rating period as a within-group factor in the ANOVA.

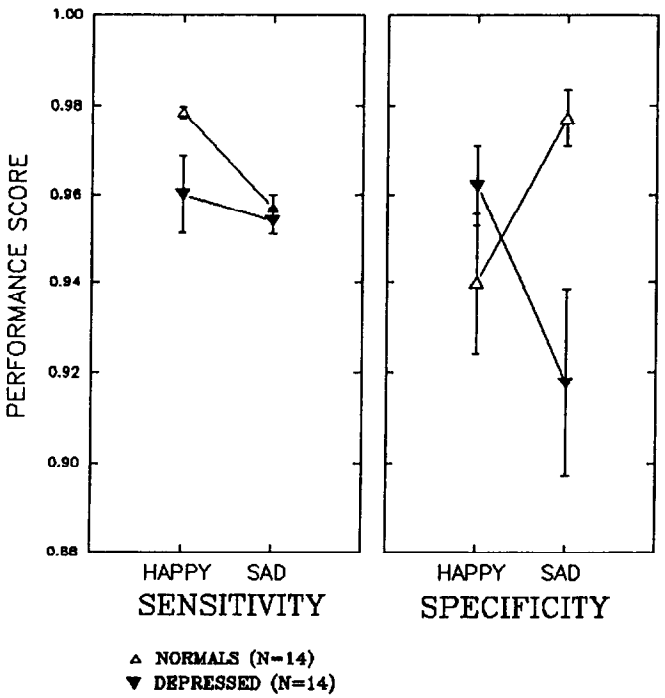
Results

Hypothesis Testing.

Hypothesis 1. The ANOVA of sensitivity scores did not yield any significant effects. Analyses of the specificity measures revealed a significant diagnosis $\times$ task interaction ($F = 4.58$; $df = 1, 26$; $p < 0.05$). Planned contrasts across groups for each task indicated that patients had lower specificity than normal controls for the sad task ($p < 0.05$, Fig. 3).

Hypothesis 2. Analyses of the negative and positive bias scores revealed a significant diagnosis $\times$ bias interaction ($F = 4.40$; $df = 1, 26$; $p < 0.05$). A planned

Fig. 3. Sensitivity and specificity scores (means and standard errors) of normal controls and depressive patients for happy-neutral and sad-neutral tasks



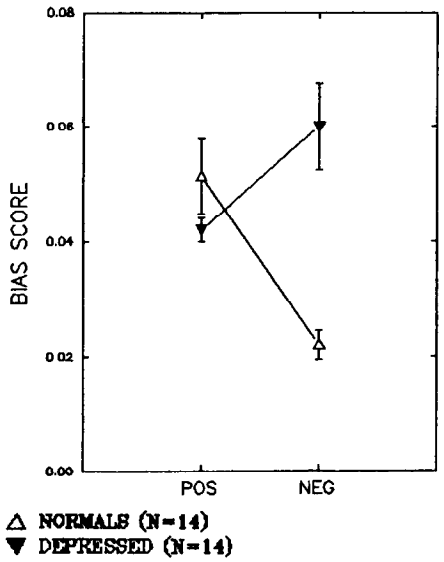
Sensitivity was defined as true positives/(true positives + false negatives). Specificity was defined as true negatives/(true negatives + false positives).

contrast across groups for negative bias indicated that patients had a greater negative bias than controls ($p < 0.01$, Fig. 4).

Hypothesis 3. PANAS trait negative mood scores correlated significantly with sensitivity for the sad discrimination task ($r = -0.71$, $p < 0.005$) in patients. Inspection of the scatterplot (Fig. 5) suggested that higher negative mood was associated with decreased sensitivity to sad discrimination in patients across the range of scores. No significant correlations were obtained in normal controls, and the correlation was significantly different between patients and controls ($z = 2.09$, $p < 0.025$). In patients, HRSD ratings correlated significantly with percent correct for sad discrimination ($r = -0.46$, $p < 0.05$) and marginally for sensitivity to sad discrimination ($r = -0.41$, $p = 0.06$).

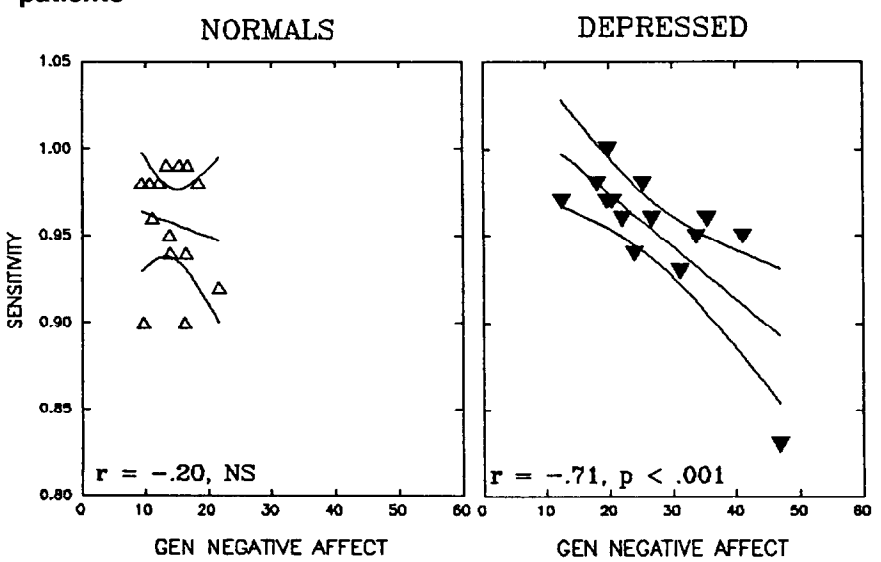
Additional Analyses. Of the six scores (happy and sad sensitivity and specificity and positive and negative bias) entered into the QROC analysis, the best discriminator between patients and controls was the negative bias measure, with a point-biserial correlation of -0.476. The best cutoff point was 0.26. Of subjects scoring higher than this, 71% were depressed, compared with only 18% of those scoring less, yielding an odds ratio of 11.2.

Fig. 4. Positive and negative bias scores (means and standard errors) of normal controls and depressive patients



Positive bias was computed as (happy-neutral false positives/total happy-neutral attempted) + (sad-neutral false negatives/total sad-neutral attempted), and negative bias was computed as (happy-neutral false negatives/total happy-neutral attempted) + (sad-neutral false positives/total sad-neutral attempted).

Fig. 5. Scatterplots of PANAS trait negative affect vs. sensitivity during the sad-neutral discrimination task in normal controls and depressive patients

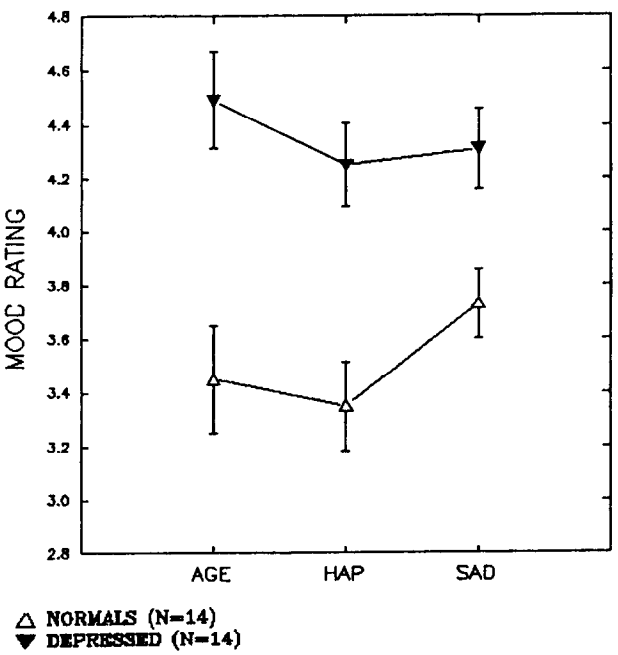


Regression lines and confidence intervals are presented for each correlation. PANAS = Positive and Negative Affect Scale.

Analysis of the percent correct data across all tasks (including age discrimination) showed a main effect for task ($F = 36.57$; $df = 2, 60$; $p < 0.001$). As in the first study on normal subjects (Erwin et al., 1992), the age-discrimination task was associated with lower performance (patients: mean = 0.81, SD = 0.03; controls: mean = 0.81, SD = 0.04) than either the happy-neutral (patients: mean = 0.94, SD = 0.02; controls: mean = 0.95, SD = 0.03) or the sad-neutral (patients: mean = 0.93, SD = 0.02; controls: mean = 0.96, SD = 0.01) emotion-discrimination tasks. No effects involving diagnosis were obtained.

The ANOVA on mood ratings obtained during task performance showed an expected main effect for diagnosis ($F = 15.47$; $df = 1, 26$; $p < 0.001$); patients had relatively more dysphoric mood than controls. There was also a task × diagnosis interaction ($F = 4.31$; $df = 2, 52$; $p < 0.025$). In normal controls, there was relatively more euphoric mood during the happy than the sad task; in patients, this effect was absent (Fig. 6).

Fig. 6. Mood ratings (means and standard errors) of normal controls and depressive patients for age discrimination, happy-neutral, and sad-neutral tasks. For orientation, a score of 3 on the vertical axis represents a slightly happy rating and a score of 5 represents a slightly sad rating.



Discussion

The present emotion-discrimination tasks showed impairment in patients with depression as compared with normal controls. This supports some of the earlier findings (see Morrison et al., 1988) in suggesting that clinical depression may affect

the processing of emotional information already at the affect-discrimination stages. Thus, patients with depression are more likely than normal controls to misinterpret the emotional valence of happy and sad facial displays.

The findings also suggest some selective deficits in depressive patients. Although the groups did not differ in overall performance as measured by the percentage of correct responses, the depressed patients showed differential deficits in measures that separate sensitivity from specificity and bias. The ability to detect neutral-sad faces was distinctively impaired, and patients showed a negative bias across happy-sad discrimination tasks. They misinterpreted neutral faces as sad and happy faces as neutral. The QROC analysis isolated the negative bias measure as the best discriminator between patients and normal controls. These findings underscore the importance of examining separately indexes of sensitivity, specificity, and bias in accordance with signal detection theory. Another study that used this methodology reported a conservative bias in patients with depression on a cognitive memory task (Corwin et al., 1990). In this study deficits in discrimination measures were observed only in severely depressed patients. Perhaps our demonstration of both a deficit in discrimination and a bias for an emotional task supports the primacy of affective disturbances in depression.

When performance was correlated with severity of negative affect, we found that patients with higher negative affect showed more impaired sensitivity to sad faces. Thus, while patients were not impaired as a group in sensitivity to sad faces, a performance deficit was evident in the more severely dysphoric patients. The present group of patients had a relatively mild and restricted range of severity of depression, patients were on medication, and the sample size was insufficient for a more comprehensive evaluation of the relationship between performance indexes and the degree of a patient's dysphoria. Conceivably, in an unmedicated sample with more severe depression, deficits in sensitivity would be more pronounced.

Analysis of the mood and anxiety scales showed some expected effects. As anticipated, patients had higher levels of trait and state anxiety than controls, although the finding that state anxiety was considerably reduced in patients suggests that engaging patients in structured activity may help ameliorate anxiety. Patients also had lower levels of positive and higher levels of negative affect than controls, as assessed by the PANAS. However, a less obvious finding is that this patient-control difference seems to reflect balanced scores on these factors in patients as compared with lopsidedly high positive and low negative affect scores in controls. This may suggest that the normal tendency is to magnify positive affect and minimize negative affect, while in depression these responses are more nearly equal. It is noteworthy that this effect is comparable to findings that patients with depression are more accurate than normal controls in assessing their own performance (Sackeim and Wegner, 1986).

Some effects were also found for mood ratings during task performance. As would be expected, patients showed more dysphoric mood than controls across tasks. However, there was only weak evidence that the emotion-discrimination tasks induced mood-congruent changes, a tendency that was observed only in normal subjects. All mood scores hovered around neutrality. Thus, emotional discrimina-

tion can apparently be measured with minimal contamination of mood-induction effects.

In summary, when a standardized emotion-discrimination task with established psychometric properties is administered to depressive patients and normal controls, an abnormality can be documented in patients. This abnormality is primarily characterized by less specificity for sad faces and a higher negative bias. Thus, although the clinical diagnosis of depression is determined by emotional expression and assessment of internal mood state, deficits also exist in the correct perception of emotional expressions. To understand and treat the psychopathology of depression, it appears to be necessary to examine more closely the conditions that give rise to misperception of emotional displays.

A limitation of the present study is the small and heterogeneous sample of patients with affective illness, including a large imbalance in sex composition. Future studies could examine sex differences, whether impaired affect recognition occurs more frequently in subtypes of depression, and whether such impairment changes with mood state or treatment and has prognostic value. For example, Corwin et al. (1990) found that a liberal ("yea-saying") bias could be observed in the manic phase and also that the bias measures were affected by pharmacologic intervention. Subsequent studies can examine the effects of emotion-discrimination tasks on regional brain activity. Such studies will determine the value of the tasks as "neurobehavioral probes" (Gur et al., 1992).

Acknowledgments. We thank Maureen Mawhinney-Hee, R. E.E.G. T., Jeff Richards, B.A., Edward Moss, Ph.D., Barbara Malamut, Ph.D., Margaret Taleff, M.S., and Joseph Smailis, B.A., for their assistance in phases of this project and the Affective Disorders Program for patient referrals. We also thank David Kupfer, M.D., for guidance on this project. This research was supported in part by the MacArthur Foundation Mental Health Research Network I and by grants MH-43880, MH-00586, and MH-45113 from the National Institute of Mental Health.

References

American Psychiatric Association. *DSM-III-R: Diagnostic and Statistical Manual of Mental Disorders.* 3rd ed., revised. Washington, DC: American Psychiatric Press, 1987.

Beck, A.T.; Ward, C.H.; Mendelson, M.; Mock, J.; and Erbaugh, J. An inventory for measuring depression. *Archives of General Psychiatry,* 4:561-571, 1961.

Cohen, J. *Statistical Power Analysis for the Behavioral Sciences.* 2nd ed. Hillside, NJ: Lawrence Erlbaum Associates, 1988.

Corwin, J.; Peselow, E.; Feenan, K.; Rotrosen, J.; Fieve, R. Disorders of decision in affective disease: An effect of β-adrenergic dysfunction? *Biological Psychiatry,* 27:813-833, 1990.

Cutting, J.C. Judgement of emotional expression in schizophrenia. *British Journal of Psychiatry,* 139:1-6, 1981.

Erwin, R.J.; Gur, R.C.; Gur, R.E.; Skolnick, B.; Mawhinney-Hee, M.; and Smailis, J. Facial emotion discrimination: I. Task construction and behavioral findings in normal subjects. *Psychiatry Research,* 42:231-240, 1992.

Feinberg, T.E.; Rifkin, A.; Schaffer, C.; and Walker, E. Facial discrimination and emotional recognition in schizophrenia and affective disorders. *Archives of General Psychiatry,* 43:276-279, 1986.

Gessler, S.; Cutting, J.; Frith, C.D.; and Weinman, J. Schizophrenic inability to judge facial emotion: A controlled study. *British Journal of Clinical Psychology,* 28:19-29, 1989.

Gur, R.C.; Erwin, R.J.; and Gur, R.E. Neurobehavioral probes for physiologic neuroimaging studies. *Archives of General Psychiatry,* 49:409-414, 1992.

Hamilton, M. A rating scale for depression. *Journal of Neurology, Neurosurgery, and Psychiatry,* 23:56-62, 1960.

Kraemer, H. Assessment of 2 × 2 associations: Generalization of signal-detection methodology. *American Statistician,* 42:37-49, 1988.

Morrison, R.L.; Bellack, A.S.; and Mueser, K.T. Deficits in facial-affect recognition and schizophrenia. *Schizophrenia Bulletin,* 14:67-83, 1988.

Sackeim, H.A., and Wegner, A. Attributional patterns in depression and euthymia. *Archives of General Psychiatry,* 43:553-560, 1986.

SAS Institute. *Statistical Analysis System.* SAS, Inc., Box 800, Cary, NC, 1985.

Saykin, A.J.; Gur, R.C.; Gur, R.E.; Kester, D.B.; Stafiniak, P.; Mozley, L.H.; Robinson, L.J.; Malamut, B.L.; Watson, B.; Shtasel, D.L.; and Mozley, P.D. Normative neuropsychological test performance: I. Effects of age, education, gender and ethnicity in a young adult urban sample. Submitted for publication.

Spielberger, C.D.; Gorsuch, R.L.; and Lushene, R.E. *Manual for the State-Trait Anxiety Inventory.* Palo Alto, CA: Consulting Psychologists Press, 1970.

Zevon, M.A., and Tellegen, A. The structure of mood change: An idiographic/nomothetic analysis. *Journal of Personality and Social Psychology,* 43:111-122, 1982.

Zuroff, D.C., and Colussy, S.A. Emotion recognition in schizophrenic and depressed patients. *Journal of Clinical Psychology,* 42:411-417, 1986.