

Discriminating between changes in bias and changes in accuracy for recognition memory of emotional stimuli

REBECCA C. GRIDER AND KENNETH J. MALMBERG
University of South Florida, Tampa, Florida

A debate has emerged as to whether recognition of emotional stimuli is more accurate or more biased than recognition of nonemotional stimuli. Teasing apart changes in accuracy versus changes in bias requires a measurement model. However, different models have been adopted by different researchers, and this has contributed to the current debate. In this article, different measurement models are discussed, and the signal detection model that is most appropriate for recognition is adopted to investigate the effects of valence and arousal on recognition memory performance, using receiver operating characteristic analyses. In addition, complementary two-alternative forced choice experiments were conducted in order to generalize the empirical findings and interpret them under a relatively relaxed set of measurement assumptions. Across all experiments, accuracy was greater for highly valenced stimuli and stimuli with high arousal value. In addition, a bias to endorse positively valenced stimuli was observed. These results are discussed within an adaptive memory framework that assumes that emotion plays an important role in the allocation of attentional resources.

Understanding the interaction between emotion and memory is central to understanding our motivations, our behavior, and our well being. In laboratory research, a study-test procedure is often used that varies the emotional content of studied materials. Such materials usually consist of words or pictures that vary in the responses they provoke in the average individual (Lang, Bradley, & Cuthbert, 1992, 2005). These normatively emotional stimuli vary on two distinct dimensions: *valence*, or the item's subjective emotional value, ranging from positive to negative; and *arousal*, or how subjectively intense or exciting it is, ranging from low to high. Although the two variables are positively correlated, a large number of prior findings have shown that both are involved in enhancing memory (e.g., Hamann, Cahill, & Squire, 1997; see Hamann, 2001, for a review).

Enhanced memory of emotional events has been attributed to the activation of the amygdala (Windmann & Kutas, 2001), the capture of attention (Cahill, Babinsky, Markowitsch, & McGaugh, 1995; Cahill & McGaugh, 1998; Estes & Adelman, in press; Hamann, 2001; Ochsner, 2000), the binding of emotional stimuli to context (MacKay et al., 2004; Mather & Nesmith, 2008), and stimulus distinctiveness (Ochsner, 2000; reviewed in Christianson, 1992). Even though one might question whether these explanations are mutually exclusive, all of them predict that memory should be better for events that elicit heightened states of emotion or arousal. However, the findings concerning enhanced memory for emotional stimuli have

come largely from free recall experiments (e.g., Danion, Kauffmann-Muller, Grange, Zimmermann, & Greth, 1995; Doerksen & Shimamura, 2001; Guy & Cahill, 1999; Hertel & Parks, 2002; Kensinger, Brierley, Medford, Growdon, & Corkin, 2002; Kensinger & Corkin, 2003). The findings in the episodic recognition literature are far less consistent in their support of these hypotheses. Some have shown that negative words are recognized better than neutral words (Comblain, D'Argembeau, Van der Linden, & Aldenhoff, 2004; Hamann, 2001; Kensinger & Corkin, 2003; Ochsner, 2000; Pesta, Murphy, & Sanders, 2001), whereas others have shown no difference in accuracy for neutral versus negative words (Doerksen & Shimamura, 2001; Ochsner, 2000). Still others have found decreased accuracy for negative stimuli (Danion et al., 1995; Dougal & Rotello, 2007; Maratos, Allan, & Rugg, 2000). In addition, some have shown that positive stimuli are recognized no better than neutral stimuli (Comblain et al., 2004; Dougal & Rotello, 2007; Ochsner, 2000), whereas others have shown that positive stimuli are recognized worse than neutral stimuli (Dougal & Rotello, 2007). Thus, there is no consensus about whether emotional stimuli are recognized better than neutral stimuli.

Some have argued that negative stimuli simply produce a more liberal response bias than do neutral stimuli (Dougal & Rotello, 2007; Leiphart, Rosenfeld, & Gabrieli, 1993; Windmann & Kutas, 2001), and several researchers have reported greater bias associated with positive stimuli (Comblain et al., 2004; Henriques, Glowacki, & David-

K. J. Malmberg, malmberg@cas.usf.edu

son, 1994; Ochsner, 2000). The literature on recognition memory for arousing material is somewhat clearer. Stimuli with a high arousal value are recognized better (Hamann et al., 1997; Kensinger & Corkin, 2003), although there is some evidence that suggests that the relationship is a non-monotonic one (Ochsner, 2000). Interestingly, although reports of bias for high-arousal words do exist (Comblain et al., 2004; Ochsner, 2000), arousal is not hypothesized to affect bias, suggesting that valance and arousal have qualitatively different effects on recognition performance.

In this article, we explore the effects of emotion and arousal on recognition memory accuracy and bias, utilizing two different methods. Our goal is to establish whether and, if so, how recognition is affected by emotional stimuli, and we hope to draw relatively strong conclusions. Toward this end, in Experiments 1 and 2, *receiver operating characteristic* (ROC) analyses were utilized to separately measure changes in accuracy and changes in bias. In Experiments 3 and 4, *two-alternative forced choice* (2AFC) analyses were utilized, which allowed us to relax the assumptions required by signal detection theory and attempt to generalize our findings. The results from both types of experiments led to consistent conclusions. Before describing our findings, however, we will begin with a primer on signal detection analyses. This will allow us to understand some of the inconsistencies in the literature within a coherent theoretical framework.

Distinguishing Between Changes in Accuracy and Bias Using ROC Analyses

The current debate is whether recognition memory for emotional materials is more accurate or more biased than recognition memory for neutral materials (e.g., Dougal & Rotello, 2007; Ochsner, 2000; Windmann & Kutas, 2001). One way of distinguishing between changes in recognition accuracy and changes in bias is to conduct ROC analyses, which are usually based on the assumptions of signal detection theory (Green & Swets, 1966). According to the classical signal detection model, recognition is based on a continuous random variable that is often conceptualized as familiarity. That is, each item elicits a level of familiarity when memory is probed, and targets (T) are more familiar on average than foils (F) because they were recently studied. The *old/new* recognition decision is based on the comparison of the familiarity associated with the test stimulus and a subjective criterion. If the familiarity of the item exceeds the criterion it is called *old*; if it does not, the item is called *new*. Thus, $\mu_T > \mu_F$, where μ_T and μ_F refer to the mean familiarities of targets and foils, respectively, and the greater the difference is between the average familiarities of targets and foils, the greater recognition accuracy will be. The top left panel of Figure 1 illustrates the signal detection model.

If we assume that the underlying familiarity values of targets and foils are normally distributed and the variances of these distributions are equal, recognition accuracy is computed as

$$d' = \frac{\mu_T - \mu_F}{\sigma} = z(\text{HR}) - z(\text{FAR}), \quad (1)$$

where HR is the observed hit rate, FAR is the observed false alarm rate, $z(X)$ is a z score transformation of probability X , and σ is the variance of the target and foil distributions of familiarity. Note that in addition to being affected by changes in the means of the target and foil distributions, d' is also affected by changes in σ . Recognition accuracy decreases as the variances of the distributions increase.

Bias is modeled as a shift in the criterion value, and common measures of bias are also derived from the observed hit and false alarm rates. One of the most commonly used measures of bias is c :

$$c = -0.5 [z(\text{HR}) + z(\text{FAR})]. \quad (2)$$

Thus, c increases as the sum of the HR and the FAR increases.

Much of the literature on the effect of emotion on recognition memory has utilized d' and c to distinguish between changes in accuracy and changes in bias (e.g., Dougal & Rotello, 2007; Ochsner, 2000). However, one of the critical assumptions underlying d' and c is disconfirmed by ROC analyses of recognition data. To see why, consider that an ROC plots the entire range of possible HR–FAR pairs for a given level of accuracy as a function of bias. This range, according to signal detection theory, is always from (0, 0) to (1, 1), not inclusive. To construct an ROC, one must, therefore, obtain several HR–FAR pairs that correspond to different levels of bias (Green & Swets, 1966; Macmillan & Creelman, 2005; Malmberg, 2002; Malmberg & Xu, 2006). This is most often accomplished using a confidence ratings task (Egan, 1958; Macmillan & Creelman, 2005; Malmberg, 2002). It is assumed that the confidence rating assigned by the subject to a given stimulus is the rating associated with the highest criterion that the familiarity of the stimulus exceeds. Thus, a different HR–FAR pair is obtained from each rating's criterion. The function relating them is the ROC, and an n -point rating scale produces an $n - 1$ point ROC. This is illustrated in the top left panel of Figure 1.

Our present assumptions make specific predictions about the form of the ratings ROC. It is curvilinear and symmetrical. This is illustrated in the top middle panel of Figure 1. When the ROC is z -transformed, a z ROC is obtained, which is shown in top right panel of Figure 1. The z ROC is linear with a slope of 1.0. Note that d' equals the distance between the z ROC and the main diagonal, which runs from $\{0, 0\}$ to $\{1, 1\}$. Thus, HR–FAR pairs that lie on the same z ROC correspond to the same level of d' (i.e., level of accuracy).

The slope of the z ROC equals the ratio of the variance of the foil distribution to the variance of the target distribution (Green & Swets, 1966). We have assumed that the underlying distributions have equal variance, and this is why the slope of the z ROC plotted in Figure 1 is equal to 1.0. In contrast, a large number of experiments have consistently shown that the slope is less than 1.0 (Egan, 1958; Ratcliff, Sheu, & Gronlund, 1992; for a review, see Parks & Yonelinas, 2007). Hence, the slope of the recognition memory z ROC disconfirms the equal-variance assumption. This has an important implication. To see why, consider the bottom three panels of Figure 1. The left panel illustrates

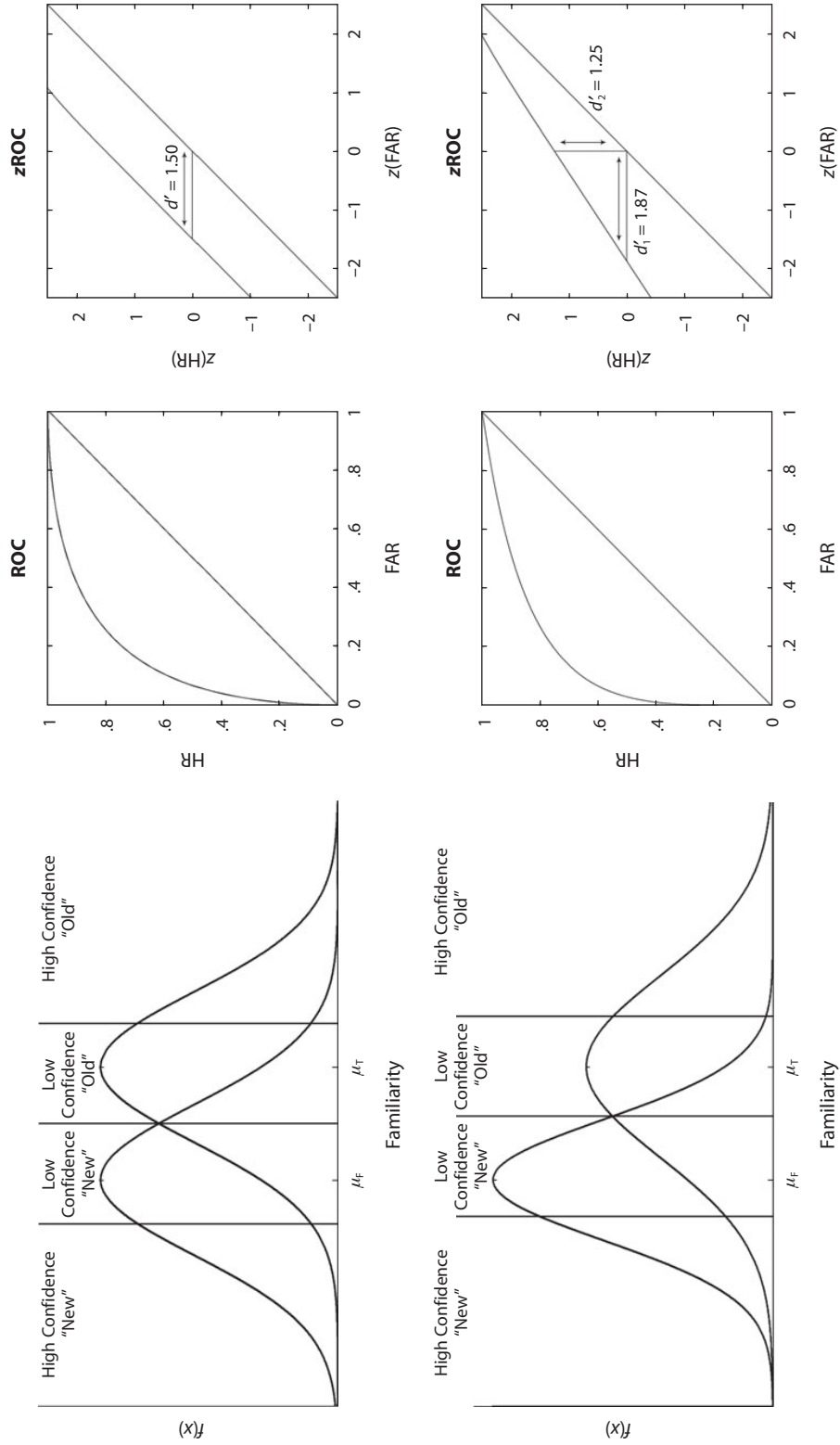


Figure 1. Equal-variance and unequal-variance signal detection models of the confidence ratings task. ROC, receiver operating characteristic; HR hit rate; FAR, false alarm rate.

the unequal-variance model that is consistent with a recognition memory zROC with a slope less than 1.0. The middle panel plots the corresponding asymmetrical ROC, and the right panel shows the corresponding zROC. The zROC no longer has a slope of 1.0 and is no longer parallel to the main diagonal, and the HR–FAR points that correspond to different levels of bias produce different values of d' (i.e., accuracy). Thus, d' and c are not independent when the equal-variance assumption is violated.

One way to address the violation of the equal-variance assumption is to obtain a multipoint zROC and to take into account its slope when computing measures of sensitivity and bias. There are four alternatives available to the researcher:

$$d'_1 = \frac{\mu_T - \mu_F}{\sigma_F} = \frac{z(\text{HR}) - z(\text{FAR})}{s}, \tag{3A}$$

$$d'_2 = \frac{\mu_T - \mu_F}{\sigma_T} = z(\text{HR}) - s\{z(\text{FAR})\}, \tag{3B}$$

$$d_a = \frac{\mu_T - \mu_F}{\text{rms}(\sigma_T, \sigma_F)} = \left(\frac{2}{1+s^2}\right)^{\frac{1}{2}} z(\text{HR}) - sz(\text{FAR}), \tag{3C}$$

and

$$d'_e = \frac{2(\mu_T - \mu_F)}{M(\sigma_T, \sigma_F)} = \frac{2}{1+s} z(\text{HR}) - sz(\text{FAR}), \tag{3D}$$

where s is the slope of the zROC and rms is the root-mean square of the variances of the underlying distributions.

There are three important things to note about this family of equations. First, all four equations are equal to d' (Equation 1) only when $s = 1.0$. Second, d'_1 and d'_2 take into account the variance of only one of the underlying distributions, whereas d_a and d'_e take into account the variances of both of the underlying distributions. Hence, we should expect that d_a and d'_e are better measures of accuracy, and they are (Macmillan & Creelman, 2005).¹ Last, and most important for present purposes, all of the measures are affected by the slope of the zROC, and hence, if there are differences in the observed slopes of the zROCs associated with different classes of stimuli, this will affect our measures of accuracy. For each measure of accuracy, there is a corresponding measure of bias:

$$c_1 = -\left(\frac{1}{1+s}\right)[z(\text{HR}) + z(\text{FAR})], \tag{4A}$$

$$c_2 = -\left(\frac{s}{1+s}\right)[z(\text{HR}) + z(\text{FAR})], \tag{4B}$$

$$c_a = -\left\{\left(\frac{\sqrt{2s}}{\sqrt{(1+s)^2}}\right)(1+s)\right\}[z(\text{HR}) + z(\text{FAR})], \tag{4C}$$

and

$$c_e = -\left(\frac{2s}{(1+s)^2}\right)[z(\text{HR}) + z(\text{FAR})]. \tag{4D}$$

Again, note that each measure of bias is affected by the slope of the zROC.

Which of the unequal-variance measures are most appropriate for recognition memory? The best measures of accuracy will minimize the differences in accuracy arising from the slope of the zROC, since accuracy should be related only to the degree to which the underlying distributions overlap, and not to the relative variances of the underlying distributions. To see why, consider Figure 2. In the top panel, the variance of the target distribution is greater than variance of the foil distribution, and the opposite situation is shown in the middle panel. In both cases, the distributions overlap by exactly the same amount, and hence, targets and foils are equally discriminable in both cases. However, the slope of the zROC obtained from the distributions in the top

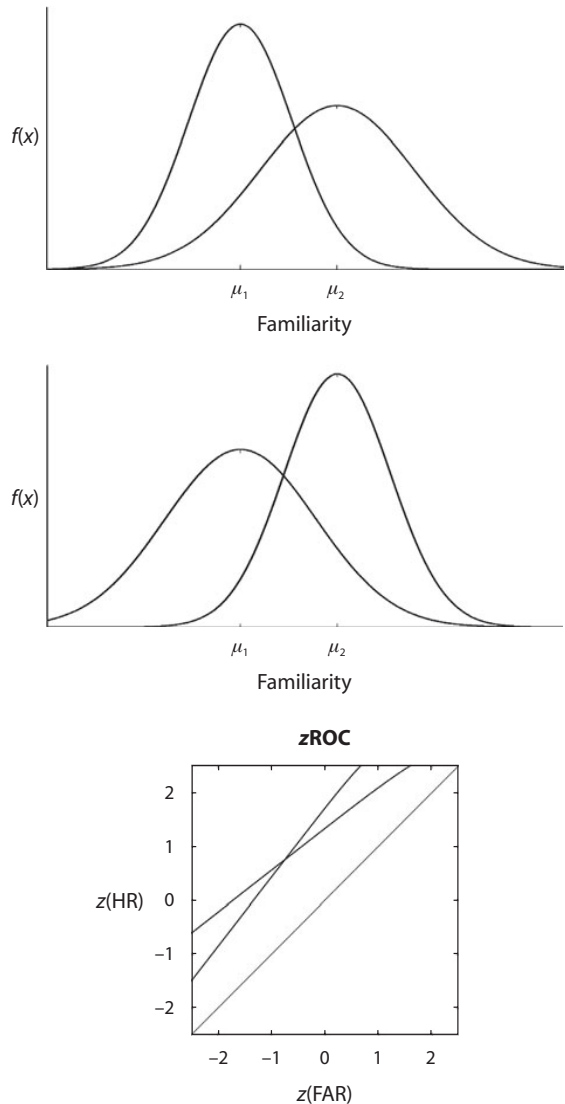


Figure 2. Signal detection models that illustrate why the relative variances of the underlying distributions should not affect accuracy. ROC, receiver operating characteristic; HR, hit rate; FAR, false alarm rate.

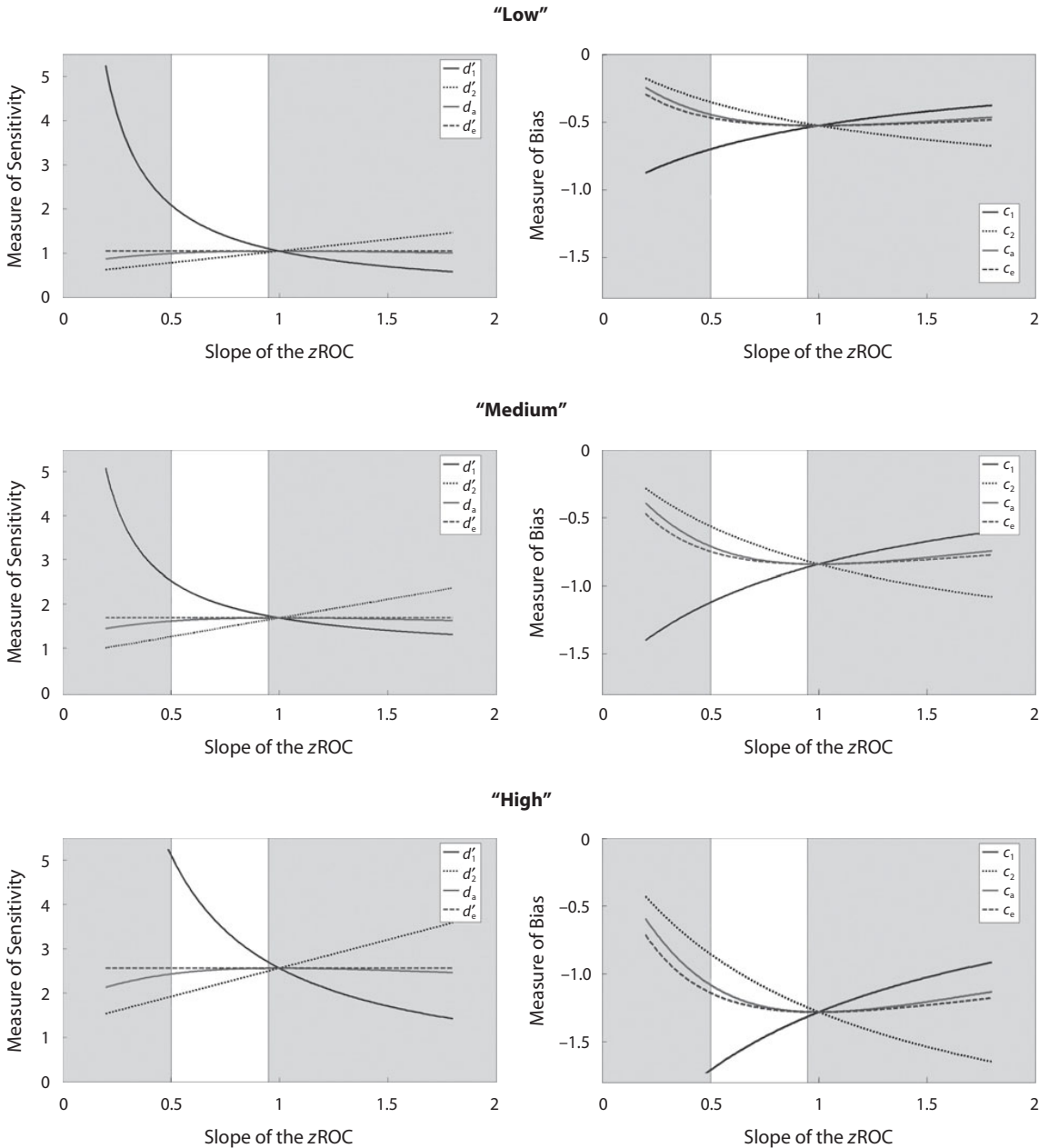


Figure 3. Theoretical dependence of receiver operating characteristic (ROC) measures of accuracy upon the slope of the ROC. Hit and false alarm rates of (.7, .3), (.8, .2), and (.9, .1) are assumed for the top, middle, and bottom panels, respectively. When the ROC slope is less than one or the variance of the *old* distribution is greater than that of the *new* distribution, measures of accuracy change as a function of the slope of the ROC. Accuracy is over- or underrepresented depending on which measure is used, and d_a is the most appropriate measure of accuracy because it is least dependent on ROC slope.

panel is less than 1.0, but the slope of the zROC obtained from the distributions in the middle panel is greater than 1.0. Thus, the unequal-variance measures will give different results even though accuracy should be the same.

So far, we have seen that the unequal-variance measures of accuracy and bias are a potential solution to the

independence problem. As it turns out, however, significant differences in the slopes of the zROCs associated with emotional stimuli have been reported (Dougal & Rotello, 2007). Thus, since all measures of accuracy and bias are going to be affected by differences in the slope of the zROC, the two important empirical questions are

the following: (1) How are accuracy and bias affected by differences in the slope of the zROC? and (2) Which measures of accuracy and bias are least affected by differences in the slope of the zROC? The answers to these questions will determine which measures are most appropriate for addressing the question of whether recognition memory for emotional stimuli is more accurate or biased.

The left panels of Figure 3 plot the various measures of accuracy (Equations 3A–3D) and bias (Equations 4A–4D) as a function of the slope of the zROC obtained under a constant HR and FAR. None of the measures are independent of the relative variances of the underlying distributions. However, some are much more affected by the slope of the zROC than are others. Furthermore, the degree to which these measures are affected by zROC slope seems to increase as accuracy increases, as is depicted in the top, middle, and bottom panels of Figure 3. We are particularly interested in how the slope affects the measures when it is between .5 and .9, because this is the interval in which the vast majority of zROC slopes have been reported (see Rotello, Macmillan, & Reeder, 2004, for a review). Figure 3 shows that the measures that take into account both the variance of the target distribution and the variance of the foil distribution are the ones least affected by the slope of the zROC (i.e., d_a , d'_e , c_a , and c_e); hence, these measures are the most appropriate ones when the slopes of the zROCs vary across conditions.

To the best of our knowledge, only one signal detection analysis of recognition memory for emotional stimuli has been reported (Dougal & Rotello, 2007). Its results suggested that emotional stimuli were recognized worse than neutral stimuli and that there was bias to respond *old* to emotional stimuli. However, we note that Dougal and Rotello used the less desirable d' and d'_1 as their dependent measures of accuracy, and they chose c as their dependent measure of bias. The problem with d' and c is that they do not independently measure accuracy and bias when recognition memory is tested. An additional problem arises from d'_1 's taking into consideration only the variance of the foil distribution. The upper panels of Figure 4 plot Dougal and Rotello's observed measures of performance as a function of the slope of the zROC associated with them. There are clear negative relationships between both Dougal and Rotello's observed levels of accuracy and bias and the slope of the zROC in a given condition. We have plotted a linear regression line to better illustrate these relationships. The strong negative relationships between d'_1 and c and slope are consistent with the theoretical anomalies associated with these measures that are plotted in Figure 3. Thus, it appears that the slope of the zROCs had an undue influence on their measures. Most important, the neutral stimuli produced zROCs with the shallowest slopes, and hence, accuracy and bias in these conditions were overestimated relative to emotional stimuli conditions.

Here, we will present four experiments that addressed memory for emotional stimuli. We manipulated the valence and arousal value of the stimuli by testing lists of normed words. In Experiments 1 and 3, we investigated memory for emotional and neutral words, and in Experiments 2 and 4, we investigated memory for arousing

and nonarousing words. We conducted ROC analyses in Experiments 1 and 2, whereas Experiments 3 and 4, we distinguish between accuracy and bias effects by using a 2AFC test procedure, in order to provide converging evidence for the conclusion drawn from the first two experiments. We will discuss this method after we present the results of Experiments 1 and 2.

EXPERIMENT 1

In Experiment 1, we varied the mean subjective emotional responses to words while controlling for a large number of other variables, including arousal. To address the measurement issues, we conducted a confidence ratings experiment in order to generate the zROCs necessary to independently measure accuracy and bias and to minimize the effect of the violation of the equal-variance assumption.

Method

Subjects. Thirty undergraduates from the University of South Florida participated in exchange for course credit.

Materials and Design. The stimuli were identical to those used by Joormann, Hertel, Brozovich, and Gotlib (2005). They consisted of 108 words drawn from the Affective Norms for English Words corpus (ANEW; Bradley & Lang, 1999). Thirty-six words were operationally defined as negative, positive, or neutral in valence. Negative words had valence ratings below 4.0 (ranging from 1 for *negative* to 9 for *positive*), neutral words had ratings between 4.0 and 6.0, and positive words had ratings above 6.0. The mean ratings were 2.3, 5.4, and 7.5 for negative, neutral, and positive words, respectively, and each stimulus class controlled for arousal value, semantic interrelatedness (as measured by latent semantic analysis matrix comparison; Landauer, Foltz, & Laham, 1998), as did Dougal and Rotello (2007), normative frequency in the natural language, and word length, so that mean values on these dimensions were not significantly different between conditions.

The presentation and the collection of data were carried out on personal computers in individual sound-attenuated subject booths. Each word was randomly assigned anew for each subject to the target and foil conditions by the computer. The study list consisted of 18 negative, positive, and neutral words, for a total study list length of 54. The order of presentation was determined randomly and anew for each subject. The words on the study list served as targets during testing, and the remaining 54 words served as foils. The order of testing was randomly determined anew for each subject.

Procedure. During the study phase, the subjects were presented one word at a time in the center of a computer screen, and they were instructed to read the words silently on the computer in preparation for a later memory test, the nature of which was unspecified. Each study word was presented for 2.5 sec with an interstimulus interval of 150 msec. The study phase was followed by a 90 sec distractor task in which the subjects mentally summed digits. The test list consisted of 108 self-paced confidence rating trials. On each trial, a single test word was presented, and the subjects' task was to rate their confidence that the word had been studied. Responses were entered on the computer keyboard. Possible responses were as follows: 1 = *very confident this word was studied*, 2 = *less confident this word was studied*, 3 = *less confident this word was not studied*, and 4 = *very confident this word was not studied*. Hence, responses 1 and 2 signify *old* responses, and 3 and 4 signify *new* responses.

Results

The results are presented in Table 1. For the statistical analyses, an alpha level of .05 was used. Significance tests of unidirectional hypotheses are one-tailed paired comparison *t* tests. Of the 29 subjects, 1 was excluded

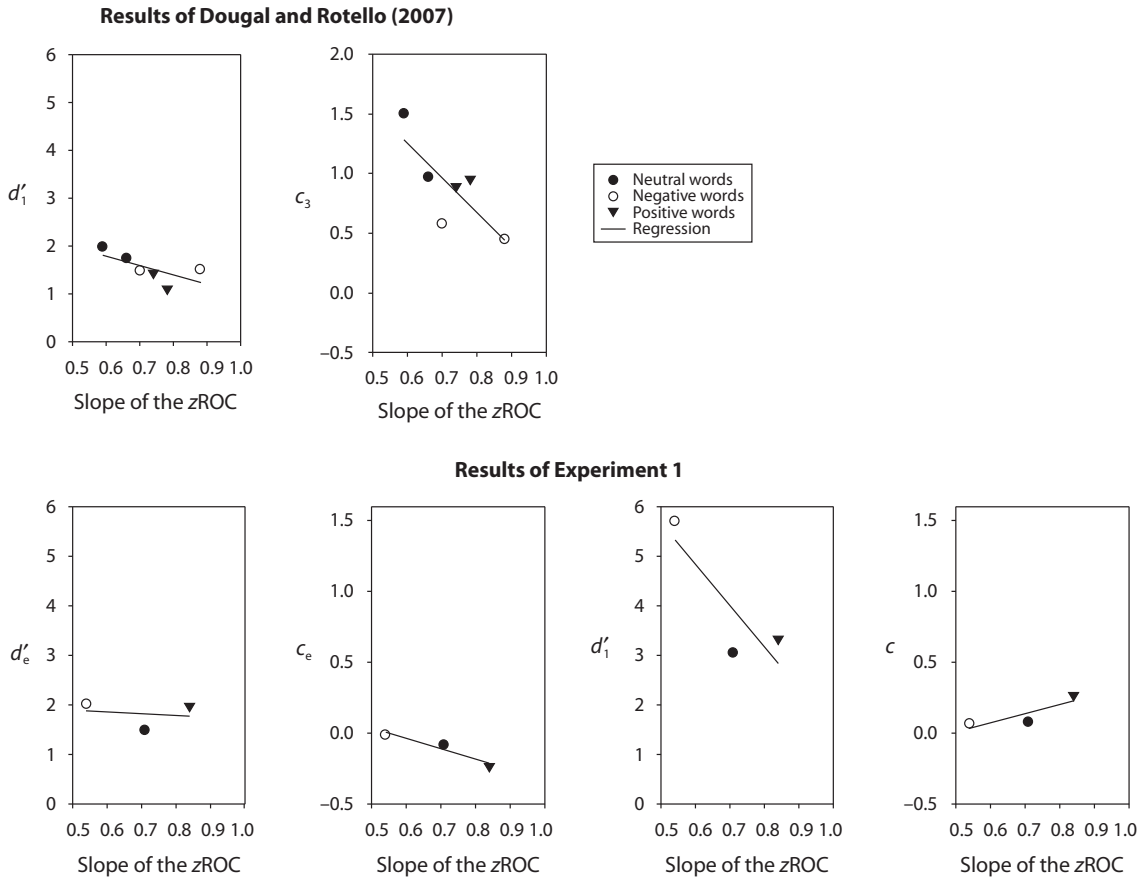


Figure 4. Reports of accuracy and bias for neutral, negative, and positive words as a function of receiver operating characteristic (ROC) slope for Dougal and Rotello (2007) in the upper panels and for Experiment 1 in the lower panels.

from analysis because the zROC produced was unusually extreme in one condition ($s = 5.65$).

zROC slopes. For all stimulus types, the slopes of the zROCs were reliably less than 1.0, suggesting that the variances of the underlying distributions were not equal. In addition, the slope of the negative-word zROC was significantly less than the slope of both the positive-word zROC [$t(27) = 1.95$] and the neutral-word zROC [$t(27) = 2.57$]. The slopes of the neutral- and positive-word zROCs were not significantly different. These findings confirm the importance of taking into account the slope of the zROC in order to minimize differences in accuracy and bias arising from differences in the relative variances of the underlying familiarity distributions between affective conditions.

Accuracy and bias. Emotional stimuli were recognized better than neutral stimuli. Accuracy as measured by d_a and d'_c was greater both for negative words [$t(27) = 1.97$ and 2.05 , respectively] and for positive words [$t(27) = 2.16$ and 2.04 , respectively] than for neutral words. There was not a reliable difference in accuracy between the negative- and the positive-word conditions. Bias as measured by c_a and c_e was more liberal for positive words than for both negative words [$t(27) = 1.98$ and

2.46 , respectively] and neutral words [$t(27) = 1.84$ and 1.73 , respectively], but there was not a reliable difference in bias for negative versus neutral words.

We next measured accuracy and bias, using d'_1 and c , the measures used by Dougal and Rotello (2007). This analysis revealed a different pattern of results, as is shown in the bottom two panels of Figure 4. There were no differences in accuracy by word condition when measured by d'_1 . However, bias as measured by c was more liberal for negative words and neutral words than for positive words [$t(27) = 1.94$ and 1.88 , respectively], and there was not a reliable difference in bias between negative and neutral words. These results demonstrate the susceptibility of the d'_1 and c measures to changes in the zROC.

EXPERIMENT 2

The results of Experiment 1 were that both negatively and positively valenced words were recognized better than neutral words and that there was a greater tendency to respond *old* to positive words than to neutral words. Another dimension of emotion is arousal. However, in Experiment 1 we controlled for arousal. The goal of Experiment 2 was to

Table 1
Results of Receiver Operating Characteristic (ROC) Analyses by
Valence and Arousal for Experiments 1 and 2

| zROC | Experiment 1 | | | | | | Experiment 2 | | | |
|--------|--------------|-----------|----------|-----------|----------|-----------|--------------|-----------|-------------|-----------|
| | Negative | | Neutral | | Positive | | High Arousal | | Low Arousal | |
| | <i>M</i> | <i>SD</i> | <i>M</i> | <i>SD</i> | <i>M</i> | <i>SD</i> | <i>M</i> | <i>SD</i> | <i>M</i> | <i>SD</i> |
| Slope | 0.54 | 0.36 | 0.71 | 0.41 | 0.84 | 0.52 | 0.80 | 0.65 | 0.83 | 0.69 |
| d_a | 1.81 | 1.20 | 1.39 | 0.73 | 1.86 | 1.21 | 1.79 | 1.04 | 1.46 | 0.81 |
| d'_e | 2.02 | 1.48 | 1.49 | 0.79 | 1.97 | 1.34 | 1.94 | 1.16 | 1.58 | 0.84 |
| c_a | -0.05 | 0.52 | -0.08 | 0.35 | -0.26 | 0.42 | -0.17 | 0.36 | -0.25 | 0.42 |
| c_e | -0.01 | 0.35 | -0.08 | 0.29 | -0.23 | 0.34 | -0.15 | 0.31 | -0.23 | 0.36 |

determine whether our findings in Experiment 1 that suggest that valenced stimuli are recognized better than neutral stimuli extend to the arousal component of emotion. Hence, arousal (high vs. low) was varied while controlling for a wide variety of other factors, including valence.

Method

Subjects, Materials, and Procedure. Sixty-three undergraduates from the University of South Florida participated in exchange for course credit. Seventy-two words were drawn from the ANEW norms to make two lists of 36 words each that were either high or low in arousal value. Mean arousal values were 5.59 for high-arousal and 3.38 for low-arousal words on a 9-point scale. We also controlled for valence (high, $M = 5.00$, $SD = 0.62$; low, $M = 5.00$, $SD = 0.37$), semantic interrelatedness (as measured by latent semantic analysis matrix comparison; Landauer et al., 1998), frequency, and word length, so that mean values were not significantly different between lists. The study list was composed of 18 randomly ordered high-arousal and 18 low-arousal words drawn from the lists of 36. The recognition procedure was identical to that in Experiment 1.

Results and Discussion

We estimated the zROC slopes using a linear regression on the observed data points. The mean slopes of the zROCs are found in Table 1. There was not a significant difference between the high- and low-arousal groups [$t(61) = 0.24$]. Paired one-tailed t test comparisons of d_a and d'_e showed greater accuracy for high-arousal words relative to low-arousal words [$t(27) = 1.97$ and 2.05 , respectively]. Although there was a slight trend suggesting a greater bias to respond *old* in the low-arousal word condition, no significant differences in c_a or c_e between high- and low-arousal words were observed [$t(61) = 1.29$ and 1.48 , respectively]. Thus, high-arousal words were recognized better than low-arousal words, but arousal did not reliably affect *old/new* response bias.

Experiments 1 and 2 produced consistent results, insofar as emotional and highly arousing stimuli produced more accurate recognition than did neutral or less arousing stimuli. In our earlier discussion, however, we demonstrated why one must take into account the slope of the zROC in order to minimize the influence of slope on our dependent measures. For instance, both d_a and d'_e are negatively related to slope (if only slightly). Hence, we questioned whether the present results might be due to the influence of the relative variances of the underlying distributions, and not to actual differences in discriminability.

To address this issue, we plotted, in the lower panels of Figure 4, our dependent measures of accuracy (d_a) and bias

(c_e) against the slopes of the zROCs that we observed in Experiment 1, in order to determine whether there is a systematic relationship between them. There were reliable differences in the slopes observed for negative words, relative to those observed for neutral or positive words. However, the regression line in the lower left panel shows very little relationship between our measure of accuracy and slope ($R^2 < .01$). Thus, it appears that the differences in accuracy associated with emotionally valenced stimuli are not due to the difference in the relative variances associated with them. Moreover, Table 1 shows that the slopes in the low-arousal and the high-arousal conditions were not significantly different, yet we observed a large difference in accuracy between these conditions. Likewise, there was a positive bias in responding only to positive words, and not to negative or neutral words. The lower right panel of Figure 4 does show a shallow negative correlation between slope and our measure of bias. Although this relationship is much smaller than the one observed in Dougal and Rotello's (2007) data, it is still possible that differences in the relative variances of the underlying distributions contributed to the reliably greater response bias that we observed for positively valenced stimuli. Since the neutral and positive words actually had the greatest slope on average, however, we would have expected that bias would have been greater in *both* of these conditions if relative variance was a major factor influencing our measure of bias, but this was not the case.

Discriminating Changes in Accuracy and Changes in Bias Using 2AFC Procedures

Emotional words were recognized better than neutral words, and there was bias to respond *old* to positively valenced stimuli. It would be desirable to generalize these findings to a different procedure, for two reasons: (1) The basis for our conclusions is a relatively large number of measurement assumptions associated with signal detection theory, and (2) even while making every effort to control for the influence of slopes on our measures of bias, we did observe a weak relationship between slope and bias. In Experiments 3 and 4, we attempted to provide converging evidence by using a 2AFC procedure.

Under standard testing conditions, the 2AFC is between one target and one foil. Bias does not enter into the equation, because both items are of the same emotional status. Thus, a relatively pure measure of accuracy is obtained by simply observing the proportion of correct choices made. Using the 2AFC procedure, for instance, Sharot and

Phelps (2004) reported greater accuracy for high-arousal words than for neutral words but did not measure differences in bias. However, the 2AFC procedure can be used to assess response bias by using null-choice comparisons (e.g., Glanzer & Adams, 1990; Glanzer, Adams, Iverson, & Kim, 1993; Hicks & Marsh, 1998; Malmberg & Murnane, 2002; Ratcliff & McKoon, 1995; Zeelenberg, Wagenmakers, & Rotteveel, 2006). For the null-choice trials, either two targets or two foils are presented that vary with respect to the factor that is manipulated experimentally. These trials are intermixed with the standard trials, and hence, the subject must simply choose which of the two alternatives was studied. Of course, there is no single correct answer for the null comparisons, and bias is revealed if there is a greater-than-chance tendency to choose one level of the factor over the other level of the factor for both the null-target comparisons and the null-foil comparisons. For instance, a null-target comparison might include a negative target and a neutral target, and a null foil comparison might include a negative foil and a neutral foil. If there is greater bias to call negative items *old*, subjects should select the negative alternatives at a greater-than-chance rate.

In the absence of response bias and in the presence of differences in accuracy, performance on the null comparison trials can tell us something important about why accuracy changed. To see why, assume that we have observed in the standard comparisons greater recognition accuracy for negative words than for neutral words. If the negative alternative is chosen less often in the null-target comparisons, but not in the null-foil comparisons, the average familiarity of the negative targets is greater than the average familiarity of neutral targets, and the average familiarity of negative and neutral foils is about the same (represented in the top panel of Figure 5). If the negative alternative is chosen less often in the null-foil comparisons, but not in the null-target comparisons, the average level of familiarity of the negative foils is less than the average familiarity of the neutral foils, and the average familiarity of negative and neutral targets is about the same (represented in the bottom panel of Figure 5). In both conditions, this is what we might expect, for instance, if negative words were encoded better or more distinctively than neutral words. In the former case, both the negative and the neutral items begin at the same baseline level of familiarity, but negative items are encoded better and, hence, seem more familiar than neutral items. In the latter case, negative foils begin at a lower baseline level of familiarity, and the additional encoding or more distinctive encoding makes them equally familiar as neutral targets. We will discuss the implications of our findings for models of memory in greater depth later. For now, we simply seek further evidence from the 2AFC procedure concerning whether emotional stimuli are recognized better than neutral stimuli and/or whether response bias is influenced by the emotional status of words.

EXPERIMENT 3

In Experiment 3, we used the 2AFC procedure in an attempt to replicate and generalize the findings from Experiment 1. Thus, we varied the valence associated with the word

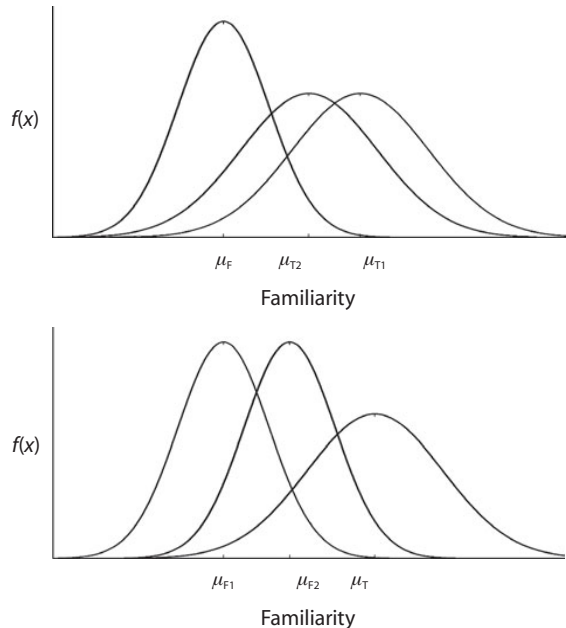


Figure 5. Signal detection models representing differences in target or foil familiarity. A greater tendency to choose the negative alternative in a null-target comparison, but not in a null-foil comparison, would indicate greater emotional target familiarity, pictured in the top panel. A lesser tendency to choose the negative alternative in a null-foil comparison, but not in a null-target comparison, would indicate lesser emotional foil familiarity, pictured in the bottom panel.

stimuli and held a large number of other variables constant, including arousal. Responses to the standard pairs allowed us to infer differences in accuracy attributed to valence. For these pairs, memory was tested with an old and a new word of the same valence. For null pairs, both words of the test pair were either old or new, and an emotional word was always paired with a neutral word. We can infer the presence of bias if, on average, subjects tend to endorse emotional words as old more often, relative to neutral words, in both the null-target pairs and the null-foil pairs.

Method

Subjects, Materials, and Procedure. Eighty undergraduates from the University of South Florida participated in exchange for course credit. Word lists were identical to those used in Experiment 1, and the study list was composed of 14 negative, 14 positive, and 18 neutral words, for a total of 46 words. The study list composition was necessary given the constraints of the 2AFC task, and as before, these words were drawn from each word list and randomly ordered to form the study list. The study procedure was identical to that in Experiment 1 and was followed by a 90-sec distractor task. The subjects next completed the 2AFC task. On each test trial, a standard or null pair was presented, and the subjects chose which word they felt had been studied. Of 46 test trials, 30 were standard pairs and 16 were null pairs.

Results and Discussion

The arc-sines of the mean probabilities of correct response scores were analyzed. Paired *t* test comparisons

of the proportions correct for the standard pairs revealed greater accuracy, relative to neutral words, for both negative words [$t(79) = 2.24, .799$ vs. $.757$] and positive words [$t(79) = 2.03, .787$ vs. $.757$]. The results of the null comparison trials are shown in Figure 6. Bias is indicated if subjects are more likely to call an emotional word *old* both in target and in foil null comparisons. One-sample t tests indicated that both negative targets [$t(79) = 4.12$] and positive targets [$t(79) = 1.92$] were more likely to be called *old* than were the neutral targets. In addition, the null comparison revealed a significant bias to respond *old* to positively valenced stimuli [$t(79) = 2.00$]. The probability of choosing a negative foil over a neutral foil did not differ reliably from $.50$. These results suggest that negative and positive targets were more familiar, on average, than neutral targets, and hence, valenced words were recognized better than neutral words. In addition, the null comparison revealed a significant bias to respond *old* to positively valenced stimuli. These findings extend our conclusions based on the findings from the ratings task to the 2AFC task used here.

EXPERIMENT 4

The results of Experiment 2 showed that high-arousal words were recognized better than low-arousal words, and the two conditions produced only a small, unreliable shift in response bias. This experiment is also a replication of Experiment 3, insofar as arousal was varied instead of valence, using the 2AFC procedure. Hence, we expected to observe better 2AFC accuracy for high-arousal words in the standard comparison trials. The null comparison trials pairing either high- and low-arousal targets or high- and low-arousal foils will allow us to further assess whether there is a bias to choose low-arousal words over high-arousal words.

Method

Subjects, Design, Materials, and Procedure. Forty-four undergraduates from the University of South Florida participated in exchange for course credit. The word lists were identical to those used in Experiment 2, and the study list was composed of 18 high-arousal and 18 low-arousal words, for a total of 36 words. These words were drawn from each word list and were randomly ordered to form the study list. The study procedure was identical to that in Experiment 3 and was followed by a 90-sec distractor task. The subjects next completed the 2AFC test phase. On each test trial, the subjects chose the word they felt had been presented previously. Of 36 test trials, 24 were standard pairs, and 12 were null pairs. The design of the test phase was similar to that in Experiment 3. The exception is that the standard pairs were composed of one target and one foil that were of similar arousal value (i.e., *high-old-high-new*, *low-old-low-new*), and the null pairs were composed of a high- and a low-arousal word, both of which were targets or foils (i.e., *high-old-low-old*, *high-new-low-new*). The null-comparison trials pairing either high- and low-arousal targets or high- and low-arousal foils allowed us to further assess whether there is a bias to choose low-arousal words over high-arousal words.

Results and Discussion

All the analyses were identical to those used in Experiment 3. Recognition accuracy was greater for high-arousal words than for low-arousal words [$.83$ vs. $.79$; $t(43) = 1.72$]. This replicates the findings from Experiment 2 that show greater recognition accuracy for high- versus low-arousal words and extend them to the 2AFC task. There was also no evidence that the arousal manipulation affected response bias. For the null comparisons, Figure 7 shows that there was a small but reliable tendency for the subjects to choose a low-arousal word over a high-arousal word when both were foils [$t(43) = 1.88$], but not when they were targets. The greater null comparison tendency to choose low-arousal foils over high-arousal foils suggests that the baseline familiarity of the low-arousal words is somewhat greater than the baseline familiarity of the high-

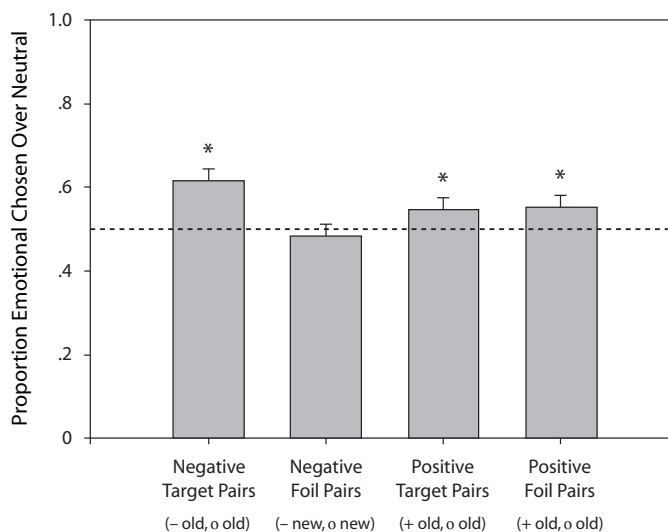


Figure 6. Proportions of emotional words chosen in null pairs by valence and novelty. Response rates above chance (.5) for negative null pairs reflect greater accuracy for negative words. Error bars display the standard errors.

arousal words, and this is the one factor that is driving the enhanced accuracy in the high-arousal condition.

GENERAL DISCUSSION

With four experiments, we investigated whether recognition memory is more accurate and/or more biased for normatively emotional words. The results were consistent for both the ROC analyses (Experiments 1 and 2) and the 2AFC analyses (Experiments 3 and 4). Emotional words were recognized better than neutral words both when the valence component of emotion was varied and when the arousal component of emotion was varied. Moreover, both emotionally negative and emotionally positive words were recognized better than neutral words. On the other hand, a greater bias to respond *old* was observed only for emotionally positive words, and not for emotionally negative words or words with high versus low arousal values.

The effects were not large, but they were consistent. Because the same pattern of results was obtained in experiments that utilized different methodologies, the present findings are generalizable, and they cannot be attributed only to a specific recognition procedure. Nor are they wed to the assumptions of the signal detection analysis of ratings data. Given the robustness of our findings, one might also ask why the literature on recognition memory of emotional words has been so unsettled. For instance, our results are often in direct conflict with those reported by Dougal and Rotello (2007), who concluded that only valence stimuli affect response bias. They based this conclusion both on signal detection estimates and on visual inspection of the ROCs. However, ROCs provide no more information than do the estimates of the criterion locations, and their estimates are not independent of sensitivity.

Another potential reason is that the levels of performance vary wildly between different reports. For instance, Windmann and Kutas (2001) reported a hit rate and a false alarm rate of about .60 and .40, respectively, for their neutral stimuli. This corresponds to a d' of only about 0.506, which is barely above chance performance. The extremely low level of accuracy in the Windmann and Kutas experiment was probably due to the extremely fast presentation rate used during study (400 msec). Under such conditions, differences in accuracy would be very difficult to detect, and bias could, therefore, play a larger role, especially when subjects are encouraged to "guess" randomly, which is what Windmann and Kutas instructed subjects to do. Their results can be compared with our results and with Ochsner's (2000) results, where accuracy was much higher. For instance, in Ochsner's experiments, recognition performance as measured by d' was, on average, approximately 1.75, which is very similar to the levels of accuracy that we report. When overall accuracy is high, it is reasonable to assume that bias will play a lesser role in recognition decisions, due to the availability of relevant episodic information obtained from memory.

Thus, it is important to maintain moderate levels of accuracy when attempting to observe changes in accuracy. This can be a challenge when investigating recognition memory, since the number of available stimuli is relatively small and,

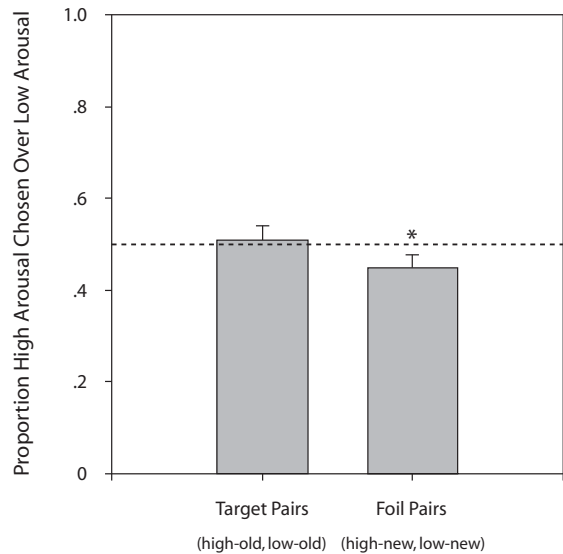


Figure 7. Proportions of high-arousal words chosen in null pairs by novelty. Response rates below chance (.5) for the foil pairs reflect greater accuracy for high-arousal words. That is, high-arousal foils are less familiar, on average, than low-arousal foils. Error bars display the standard errors.

hence, recognition accuracy is often very high. Exercising adequate control over confounding stimulus factors, such as concreteness, semantic relatedness, frequency, and word length, decreases the number of available stimuli and further compounds this issue. Importantly, the small number of available stimuli is not a problem when investigating free recall, since it is a much more difficult task. Hence, this helps to explain why free recall of emotional words is typically reported to be better than free recall of neutral words, and why memory for emotional stimuli has sometimes been reported to be task dependent.

Another potential reason for the unsettled nature of the prior literature on recognition of emotional stimuli is that different researchers utilized different measurements of accuracy and bias. In the introduction, we showed how factors unrelated to accuracy and bias can affect measures that do not take them into account. We were particularly interested in controlling for the slope of the recognition memory z ROC (i.e., the relative variances of the underlying familiarity distributions), since it had been found to vary between emotionally valenced and neutral words (Dougal & Rotello, 2007). On the basis of our analyses, we concluded that d_a and d'_e were the measures of recognition accuracy that were least susceptible to variations in slope. Likewise, we concluded that c_a and c_e were the best measures of recognition bias. When these were used as dependent measures in a confidence ratings experiment, the results consistently showed greater accuracy for emotional words.

The results from the 2AFC experiments support the same conclusions as the results from the confidence ratings experiments. They also shed some light onto the nature of memory for emotional words. We found that emotionally valenced and arousing words were recognized better than

neutral words. In addition, we found a null comparison advantage only for the emotionally valenced targets. This pattern of forced choice performance supports the inference that the average familiarity of emotional targets is greater than the average familiarity of neutral targets but that the average familiarity of emotional and neutral foils is about the same. A different picture emerged for the accuracy difference observed for high- versus low-arousal words. Although standard comparisons showed greater accuracy for high-arousal words, there was a null comparison advantage only for high- versus low-arousal foils. That is, the subjects were more likely to choose the low-arousal foil over the high-arousal foil, suggesting that the low-arousal foils were more familiar, on average, than the high-arousal foils. We were somewhat surprised by this finding, and thus we conducted a post hoc analysis of the results from Experiment 2. They showed that there was significant difference in false alarm rates (low arousal, .32; high arousal, .25), but not in hit rates (low arousal, .81; high arousal, .83) for high- versus low-arousal words. Thus, emotionally valenced words are recognized better because of differences in target familiarity, and high-arousal words are recognized better than low-arousal words because high-arousal foils are less familiar than low-arousal foils.

Although the mnemonic status of valenced versus arousing words may differ prior to the experiment, the mechanism involved in producing better memory for them may be the same. For instance, the emotional status of a stimulus could enhance mnemonic organization (e.g., Sison & Mather, 2007) and/or the binding of item and context information (MacKay et al., 2004; Mather & Nesmith, 2008). Here, however, we want to consider a simple extension of Malmberg and Murnane's (2002) retrieving effectively from memory (REM) model of the word frequency effect. Malmberg and Murnane assumed that a rational cognitive system should allocate more attention to those elements of our environment that are uncommon. In order to account for their findings that uncommon words are recognized better when they have been studied on a list dominated by common words, Malmberg and Murnane further assumed that uncommon words attract more attention when studied in contexts dominated by common words. As a result, uncommon words are encoded more accurately and, hence, more distinctively when studied in a context dominated by common words. In subsequent experiments, Malmberg and Nelson (2003; see also Criss & Shiffrin, 2005) reported evidence that supported the assumption that uncommon words attract more attention than do common words.

Within the same framework of a rational cognitive system, one could assume that attention should be allocated to stimuli on the basis of the costs and rewards associated with them and that emotional stimuli have greater costs and rewards associated with them than do nonemotional stimuli, either because of the emotions that these words evoke or because of the arousal that they evoke. Hence, emotionally positive and negative words should receive more attention than do neutral words, and high-arousal words more attention than do low-arousal words. The literature is converging on this assumption. For instance, emotional words are perceived more readily than neutral

words regardless of whether the mode of presentation is visual (Zeelenberg et al., 2006) or auditory (Weger, Meier, Robinson, & Inhoff, 2007). In addition, when two stimuli are visually presented and one stimulus is emotionally valenced and the other is not, eye-tracking data indicate that the initial eye fixation is more likely to land on the emotionally valenced stimulus (Knight et al., 2007). These findings suggest that relatively high-level information can guide attention (e.g., Sanborn, Malmberg, & Shiffrin, 2004), that stimuli high in emotional status are more salient than neutral stimuli, and that hence, the attentional system is biased to select them for processing. If so, most models would assume that the occurrence of emotional stimuli should be remembered better than the occurrence of neutral stimuli, regardless of whether memory is tested via a free recall or a recognition procedure.

Here, we want to account for our findings that suggest that emotionally valenced words are recognized better because the average familiarity of valenced targets is greater relative to the average familiarity of neutral targets *and* for those that suggest that highly arousing words are recognized better because the average familiarity of highly arousing foils is lower than the average familiarity of nonarousing foils. In order to do so, we extend Malmberg and Murnane's (2002) REM model by assuming that the preexperimental familiarity of emotionally valenced and neutral words are approximately equal. Therefore, enhanced processing of emotionally valenced stimuli increases their mean familiarity, relative to neutral stimuli. This produces better recognition memory for emotionally valenced words and a greater tendency to choose emotionally valenced targets on null comparison forced choice trials.

We further assume that the preexperimental familiarity of highly arousing words is lower, on average, than that of relatively nonarousing words. This may be because arousing words are more distinctive than nonarousing words (e.g., Malmberg, Steyvers, Stephens, & Shiffrin, 2002; Schmidt, 1991) or because they tend to occur in fewer different contexts than do nonarousing words (Dennis & Humphreys, 2001; Steyvers & Malmberg, 2003). Enhanced processing during study then increases the familiarity of highly arousing targets to a greater degree than it does that of less arousing targets, but not to the degree that would produce a greater mean familiarity of highly arousing targets, as compared with the mean familiarity of less arousing targets. This would produce better recognition memory for highly arousing words and a lesser tendency to choose highly arousing foils on null comparison forced choice trials. Last, according to the Malmberg and Murnane (2002) model, the allocation of resources during study might be related to the composition of the study lists. For instance, if a pure-list design was used instead of the mixed-list design that we used here, subjects might not selectively attend to emotional words to a greater degree than they do to unemotional words.

We also observed a greater tendency to indicate that positively valenced words were studied than to indicate that neutral words were. Comblain et al. (2004), Ochsner (2000), and Dougal and Rotello (2007) reported similar recognition findings. In addition, others have found evi-

dence supporting a liberal response bias for recall of positive flashbulb memories (Bohn & Berntsen, 2007) and a selection bias for remembering positive autobiographical events (Schulkind & Woldorf, 2005). Induction of a positive affective state also increased the recognition bias of neutral words (Phaf & Rotteveel, 2005). These positive bias effects are in line with the Pollyanna principle, which states that positive stimuli are favored over time (Matlin & Stang, 1978). Our data from the null comparison forced choice trials suggest that such positive biases are most likely to occur when relevant evidence on which to make a decision is limited and, hence, reflect preexperimental knowledge about the stimuli. Interestingly, we can report no effects on bias of high- versus low-arousal words. When combined with the forced choice results that indicate differences in target familiarity between emotionally valenced and neutral targets but differences in foil familiarity for high- versus low-arousal words, our findings support a distinction between the two traditional components of emotion: valence and arousal.

Thus, the present findings are consistent with a model that assumes that emotional words are encoded in some fashion that better supports recognition. However, it is possible that the differences in encoding also affect the manner in which these traces are retrieved. Whether recollection contributes to recognition—and if so, to what degree—is an important issue. Recollection may play a more important role for emotional words. However, the data from the ROC experiments do not seem to support this hypothesis. According to most models (e.g., Arndt & Reder, 2003; Yonelinas, 1994), there should be an inverse relationship between the slope of the zROC and the contribution of recollection. However, in Experiment 1, the accuracy was greater for both positive and negative words, but the slope of the zROC was unrelated to accuracy. Likewise, in Experiment 2, there was a large difference in accuracy for arousing versus nonarousing words, but there was no difference in the slope of the zROC.

CONCLUSIONS

Prior experiments investigating the effect of emotion on recognition memory have produced mixed results. One reason for this may be the use of inconsistent and often inappropriate dependent measures. Another reason is that overall levels of performance were sometimes very low. With four experiments that used appropriate measures but very different procedures, we found that normatively emotional stimuli are recognized better than normatively neutral stimuli, and we found that positively valenced words are more likely to elicit a positive recognition response than are neutral or negative words. The results are consistent with intuition, and they support a model of memory that assumes that emotion influences the allocation of attentional resources.

AUTHOR NOTE

Correspondence concerning this article should be addressed to K. J. Malmberg, Department of Psychology, University of South Florida, 4202 East Fowler Ave., PCD 4118G, Tampa, FL 33620 (e-mail: malmberg@cas.usf.edu).

REFERENCES

ARNDT, J., & REDER, L. M. (2003). The effect of distinctive visual information on false recognition. *Journal of Memory & Language*, *48*, 1-15.

BOHN, A., & BERNTSEN, D. (2007). Pleasantness bias in flashbulb memories: Positive and negative flashbulb memories of the fall of the Berlin Wall among East and West Germans. *Memory & Cognition*, *35*, 565-577.

BRADLEY, M. M., & LANG, P. J. (1999). *Affective norms for English words (ANEW): Stimuli, instruction manual and affective ratings* (Tech. Rep. C-1). Gainesville: University of Florida, Center for Research in Psychophysiology.

CAHILL, L., BABINSKY, R., MARKOWITSCH, H. J., & MCGAUGH, J. L. (1995). The amygdala and emotional memory. *Nature*, *377*, 295-296.

CAHILL, L., & MCGAUGH, J. L. (1998). Mechanisms of emotional arousal and lasting declarative memory. *Trends in Neurosciences*, *21*, 294-299.

CHRISTIANSON, S.-Å. (1992). Emotional stress and eyewitness memory: A critical review. *Psychological Bulletin*, *112*, 284-309.

COMBLAIN, C., D'ARGEMBEAU, A., VAN DER LINDEN, M., & ALDENHOFF, L. (2004). The effect of ageing on the recollection of emotional and neutral pictures. *Memory*, *12*, 673-684.

CRISS, A. H., & SHIFFRIN, R. M. (2005). List discrimination in associative recognition and implications for representation. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, *31*, 1199-1212.

DANION, J.-M., KAUFFMANN-MULLER, F., GRANGÉ, D., ZIMMERMANN, M.-A., & GRETH, P. (1995). Affective valence of words, explicit and implicit memory in clinical depression. *Journal of Affective Disorders*, *34*, 227-234.

DENNIS, S., & HUMPHREYS, M. S. (2001). A context noise model of episodic word recognition. *Psychological Review*, *108*, 452-478.

DOERKSEN, S., & SHIMAMURA, A. P. (2001). Source memory enhancement for emotional words. *Emotion*, *1*, 5-11.

DOUGAL, S., & ROTELLO, C. M. (2007). "Remembering" emotional words is based on response bias, not recollection. *Psychonomic Bulletin & Review*, *14*, 423-429.

EGAN, J. P. (1958). *Recognition memory and the operating characteristic* (Tech. Note AFCRC-TN-58-51). Bloomington: Indiana University, Hearing and Communication Laboratory.

ESTES, Z., & ADELMAN, J. S. (in press). Automatic vigilance for negative words in lexical decision and naming: Comment on Larsen, Mercer, and Balota (2006). *Emotion*.

GLANZER, M., & ADAMS, J. K. (1990). The mirror effect in recognition memory: Data and theory. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, *16*, 5-16.

GLANZER, M., ADAMS, J. K., IVERSON, G. J., & KIM, K. (1993). The regularities of recognition memory. *Psychological Review*, *100*, 546-567.

GREEN, D. M., & SWETS, J. A. (1966). *Signal detection theory and psychophysics*. New York: Wiley.

GUY, S. C., & CAHILL, L. (1999). The role of overt rehearsal in enhanced conscious memory for emotional events. *Consciousness & Cognition*, *8*, 114-122.

HAMANN, S. [B.] (2001). Cognitive and neural mechanisms of emotional memory. *Trends in Cognitive Sciences*, *5*, 394-400.

HAMANN, S. B., CAHILL, L., & SQUIRE, L. R. (1997). Emotional perception and memory in amnesia. *Neuropsychology*, *11*, 104-113.

HENRIQUES, J. B., GLOWACKI, J. M., & DAVIDSON, R. J. (1994). Reward fails to alter response bias in depression. *Journal of Abnormal Psychology*, *103*, 460-466.

HERTEL, P. T., & PARKS, C. (2002). Emotional episodes facilitate word recall. *Cognition & Emotion*, *16*, 685-694.

HICKS, J. L., & MARSH, R. L. (1998). A decrement-to-familiarity interpretation of the revelation effect from forced-choice tests of recognition memory. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, *24*, 1105-1120.

JOORMANN, J., HERTEL, P. T., BROZOVICH, F., & GOTLIB, I. H. (2005). Remembering the good, forgetting the bad: Intentional forgetting of emotional material in depression. *Journal of Abnormal Psychology*, *114*, 640-648.

KENSINGER, E. A., BRIERLEY, B., MEDFORD, N., GROWDON, J. H., & CORKIN, S. (2002). Effects of normal aging and Alzheimer's disease on emotional memory. *Emotion*, *2*, 118-134.

- KENSINGER, E. A., & CORKIN, S. (2003). Memory enhancement for emotional words: Are emotional words more vividly remembered than neutral words? *Memory & Cognition*, **31**, 1169-1180.
- KNIGHT, M., SEYMOUR, T. L., GAUNT, J. T., BAKER, C., NESMITH, K., & MATHER, M. (2007). Aging and goal-directed emotional attention: Distraction reverses emotional biases. *Emotion*, **7**, 705-714.
- LANDAUER, T. K., FOLTZ, P. W., & LAHAM, D. (1998). An introduction to latent semantic analysis. *Discourse Processes*, **25**, 259-284.
- LANG, P. J., BRADLEY, M. M., & CUTHBERT, B. N. (1992). A motivational analysis of emotion: Reflex-cortex connections. *Psychological Science*, **3**, 44-49.
- LANG, P. J., BRADLEY, M. M., & CUTHBERT, B. N. (2005). *International Affective Picture System (IAPS): Digitized photographs, instruction manual and affective ratings*. (Tech. Rep. A-6). Gainesville: University of Florida, Center for Research in Psychophysiology.
- LEIPHART, J., ROSENFELD, J. P., & GABRIELI, J. D. (1993). Event-related potential correlates of implicit priming and explicit memory tasks. *International Journal of Psychophysiology*, **15**, 197-206.
- MACKAY, D. G., SHAFTO, M., TAYLOR, J. K., MARIAN, D. E., ABRAMS, L., & DYER, J. R. (2004). Relations between emotion, memory, and attention: Evidence from taboo Stroop, lexical decision, and immediate memory tasks. *Memory & Cognition*, **32**, 474-488.
- MACMILLAN, N. A., & CREELMAN, C. D. (2005). *Detection theory: A user's guide* (2nd ed.). Mahwah, NJ: Erlbaum.
- MALMBERG, K. J. (2002). On the form of ROCs constructed from confidence ratings. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **28**, 380-387.
- MALMBERG, K. J., & MURNANE, K. (2002). List composition and the word-frequency effect for recognition memory. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **28**, 616-630.
- MALMBERG, K. J., & NELSON, T. O. (2003). The word frequency effect for recognition memory and the elevated-attention hypothesis. *Memory & Cognition*, **31**, 35-43.
- MALMBERG, K. J., STEYVERS, M., STEPHENS, J. D., & SHIFFRIN, R. M. (2002). Feature frequency effects in recognition memory. *Memory & Cognition*, **30**, 607-613.
- MALMBERG, K. J., & XU, J. (2006). The influence of averaging and noisy decision strategies on the recognition memory ROC. *Psychonomic Bulletin & Review*, **13**, 99-105.
- MARATOS, E. J., ALLAN, K., & RUGG, M. D. (2000). Recognition memory for emotionally negative and neutral words: An ERP study. *Neuropsychologia*, **38**, 1452-1465.
- MATHER, M., & NESMITH, K. (2008). Arousal-enhanced location memory for pictures. *Journal of Memory & Language*, **58**, 449-464.
- MATLIN, M. W., & STANG, D. J. (1978). *The Pollyanna principle: Selectivity in language, memory, and thought*. Cambridge, MA: Schenkman.
- OCHSNER, K. N. (2000). Are affective events richly recollected or simply familiar? The experience and process of recognizing feelings past. *Journal of Experimental Psychology: General*, **129**, 242-261.
- PARKS, C. M., & YONELINAS, A. P. (2007). Moving beyond pure signal-detection models: Comment on Wixted (2007). *Psychological Review*, **114**, 188-202.
- PESTA, B. J., MURPHY, M. D., & SANDERS, R. E. (2001). Are emotionally charged lures immune to false memory? *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **27**, 328-338.
- PHAF, R. H., & ROTTEVEEL, M. (2005). Affective modulation of recognition bias. *Emotion*, **5**, 309-318.
- RATCLIFF, R., & MCKOON, G. (1995). Bias in the priming of object decisions. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **21**, 754-767.
- RATCLIFF, R., SHEU, C. F., & GRONLUND, S. D. (1992). Testing global memory models using ROC curves. *Psychological Review*, **99**, 518-535.
- ROTELLO, C. M., MACMILLAN, N. A., & REEDER, J. A. (2004). Sum-difference theory of remembering and knowing: A two-dimensional signal-detection model. *Psychological Review*, **111**, 588-616.
- SANBORN, A. N., MALMBERG, K. J., & SHIFFRIN, R. M. (2004). High-level effects of masking on perceptual identification. *Vision Research*, **44**, 1427-1436.
- SCHMIDT, S. R. (1991). Can we have a distinctive theory of memory? *Memory & Cognition*, **19**, 523-542.
- SCHULKIND, M. D., & WOLDORF, G. M. (2005). Emotional organization of autobiographical memory. *Memory & Cognition*, **33**, 1025-1035.
- SHAROT, T., & PHELPS, E. A. (2004). How arousal modulates memory: Disentangling the effects of attention and retention. *Cognitive, Affective, & Behavioral Neuroscience*, **4**, 294-306.
- SISON, J. A. G., & MATHER, M. (2007). Does remembering emotional items impair recall of same-emotion items? *Psychonomic Bulletin & Review*, **14**, 282-287.
- STEVYERS, M., & MALMBERG, K. J. (2003). The effect of normative context variability on recognition memory. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **29**, 760-766.
- SWETS, J. A., & PICKETT, R. M. (1982). *Evaluation of diagnostic systems: Methods from signal detection theory*. New York: Academic Press.
- WEGER, U. W., MEIER, B. P., ROBINSON, M. D., & INHOFF, A. W. (2007). Things are sounding up: Affective influences on auditory tone perception. *Psychonomic Bulletin & Review*, **14**, 517-521.
- WINDMANN, S., & KUTAS, M. (2001). Electrophysiological correlates of emotion-induced recognition bias. *Journal of Cognitive Neuroscience*, **13**, 577-592.
- YONELINAS, A. P. (1994). Receiver-operating characteristics in recognition memory: Evidence for a dual-process model. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **20**, 1341-1354.
- ZEELLENBERG, R., WAGENMAKERS, E.-J., & ROTTEVEEL, M. (2006). The impact of emotion on perception: Bias or enhanced processing? *Psychological Science*, **17**, 287-291.

NOTE

1. Although A' does not measure sensitivity independently of bias, d_a does, and A_z is a simple transformation of d_a ; thus, one needs to compute d_a in order to compute A_z . To compute A_z directly from the ROC, one needs a computer fitting algorithm (Swets & Pickett, 1982). Hence, we prefer d_a .

(Manuscript received December 16, 2007;
revision accepted for publication February 13, 2008.)