Pragmatics of Measuring Recognition Memory: Applications to Dementia and Amnesia

Joan Gay Snodgrass
New York University

June Corwin
New York University School of Medicine and New York
Veterans Administration Medical Center

SUMMARY

This article has two purposes. The first is to describe four theoretical models of yes–no recognition memory and present their associated measures of discrimination and response bias. These models are then applied to a set of data from normal subjects to determine which pairs of discrimination and bias indices show independence between discrimination and bias. The following models demonstrated independence: a two-high-threshold model, a signal detection model with normal distributions using d' and C (rather than beta), and a signal detection model with logistic distributions and a bias measure analogous to C. C is defined as the distance of criterion from the intersection of the two underlying distributions.

The second purpose is to use the indices from the acceptable models to characterize recognition memory deficits in dementia and amnesia. Young normal subjects, Alzheimer's disease patients, and parkinsonian dementia patients were tested with picture recognition tasks with repeated study–test trials. Huntington's disease patients, mixed etiology amnesics, and age-matched normals were tested by Butters, Wolfe, Martone, Granholm, and Cermak (1985) using the same paradigm with word stimuli. Demented and amnesic patients produced distinctly different patterns of abnormal memory performance. Both groups of demented patients showed poor discrimination and abnormally liberal response bias for words (Huntington's disease) and pictures (Alzheimer's disease and parkinsonian dementia), whereas the amnesic patients showed the worst discrimination but normal response bias for words. Although both signal detection theory and two-high-threshold discrimination parameters showed identical results, the bias measure from the two-high-threshold model was more sensitive to change than the bias measure (C) from signal detection theory.

Three major points are emphasized. First, any index of recognition memory performance assumes an underlying model. Second, even acceptable models can lead to different conclusions about patterns of learning and forgetting. Third, efforts to characterize and ameliorate abnormal memory should address both discrimination and bias deficits.

A common method for measuring memory performance is by use of a recognition memory test. Here we consider the yes–no form of recognition memory test in which studied (old) items are presented mixed with distractor (new) items and the subject's task is to decide whether each item is old (by responding "yes") or new (by responding "no"). Performance on a yes–no recognition test is summarized by two measures: the hit rate, the probability that the subject classifies an old item as old, and the false alarm rate, the probability that the subject classifies a new item as old. A given model of the recognition memory process specifies how these two measures are to be combined to characterize memory performance.

Since Banks (1970) and Lockhart and Murdock (1970) reviewed the use of signal detection theory (SDT) in measurement of human memory, researchers in the clinical areas have conscientiously adopted the indices derived from SDT to evaluate memory in clinical populations (e.g., Branconnier, Cole, Spera, & DeVitt, 1982; Butters, Wolfe, Martone, Granholm, &

Cermak, 1985; Mohs & Davis, 1982). In contrast, researchers investigating normal memory often use a simpler measure of discrimination such as hits minus false alarms or the nonparametric measure A' (e.g., Gillund & Shiffrin, 1984; Tulving & Thomson, 1971).

Much of this literature in both normal and abnormal populations focuses on evaluation of discrimination with relatively little attention paid to the other half of the recognition story, response bias. However, in the abnormal literature it is becoming increasingly clear that abnormal bias is an important component of abnormal memory (Branconnier et al., 1982; Butters et al., 1985). In fact, Branconnier et al. demonstrated that false alarm rate fared almost as well as d' in discriminating between Alzheimer's disease and normal elderly memory.

In this article, we reexamine the models underlying the use of various indices of recognition memory, and make some suggestions for appropriately measuring both bias and discrimination in normal and abnormal populations. The basic desiderata

of such indices are twofold; first, that the discrimination index be invariant across explicit manipulations of bias; and second, that the bias index be invariant across explicit manipulations of discriminability. Otherwise, comparisons between populations having different discrimination performance may on the one hand yield spurious differences in bias, or on the other fail to detect differences in bias which are present. We also note that Murdock (1974, pp. 34–35) proposed and tested a similar criterion for evaluating the signal detection model of recognition memory.

We will first describe four models of the recognition memory process and their associated indices. A critical feature of any model of recognition memory is that it defines a potentially infinite set of hit/false alarm rate pairs that yield equivalent discrimination across all levels of bias. These equal discrimination functions have been called isomemory functions or memory-operating characteristics. We will use the term *isomemory function* in the remainder of this article. In addition, each model defines a set of isobias functions, representing sets of hit/false alarm rate pairs that yield equivalent bias across all levels of discrimination.

After describing the isomemory and isobias functions for each model, we next apply these four models to data from an experiment designed to independently manipulate discrimination and bias. The degree to which a model's discrimination and bias indices show the desired invariances will determine the acceptability of the indices (and therefore the model).

Finally, we will use the acceptable indices to characterize learning and forgetting of demented, amnesic, and normal subjects in repeated study–test tasks with picture and word stimuli.

The Recognition Memory Task

All indices of recognition memory performance considered here are based on a single pair of hit and false alarm rates, as would be obtained in a yes–no recognition memory test procedure. In conformity with common usage, we define a hit as a yes response to an old item, and the hit rate, H, as the conditional probability of responding yes to an old item, P(yes/old). A false alarm is defined as a yes response to a new item, and the false alarm rate, FA, as the conditional probability of responding yes to a new item, P (yes/new). Similarly, a correct rejection is de-

Figure 1. Stimulus–response matrix (A) and corrected matrix (B) illustrating the correction for computing hit and false alarm rates.

fined as a no response to a new item, and the correct rejection rate, CR, as the conditional probability of responding no to a new item, P (no/new). A miss is defined as a no response to an old item, and the miss rate, M, as the conditional probability of responding no to an old item, P (no/old). Because the sum of the hit and miss rates is 1.0, the hit rate is sufficient to describe what happened on old item trials. Similarly, because the sum of the false alarm and correct rejection rates is also 1.0, the false alarm rate is sufficient to describe what happened on new item trials. The pair of hit and false alarm rates thus completely summarizes the data of a single subject in a single condition of a yes–no recognition memory test.

Measures for the two SDT models are undefined for hit rates of 1.0 or false alarm rates of 0 because the corresponding z scores are infinite. Accordingly, we have corrected all hit and false alarm rates by adding 0.5 to each frequency and dividing by $N + 1$, where N is the number of old or new trials. This correction is recommended for log-linear models (e.g., Upton, 1978). For consistency, we recommend applying the correction routinely, even in the absence of 0s and 1s and even when SDT measures are not calculated. Figure 1 presents the stimulus–response matrix with the correction shown. This transformation has been applied to all the data reported here.

The Four Models of Recognition Memory

Signal Detection Theory

The signal detection model for recognition memory proposes that items presented for test in a recognition memory task lie along a continuum of familiarity or memory strength. Some old items will have high familiarity, most will have medium familiarity, and some will have low familiarity. New items (distractors) will have similarly distributed levels of familiarity but with a lower mean level than the old items. Theoretically, the distributions of familiarity of old and new items will always overlap. The distribution of familiarity for old items corresponds to the signal + noise distribution of sensory SDT, and the distribution of familiarity for new items corresponds to the noise alone distribution of sensory SDT.

It is assumed that subjects cannot directly determine whether or not an item is old. Rather, the subject is aware only of an item's familiarity or strength value. Thus, the subject must set some criterion value of familiarity, called x_c, such that items whose familiarity exceeds the criterion are called "old" and

those that fail to exceed the criterion are called "new." The proportion of old items exceeding the criterion is estimated by the hit rate, and the proportion of new items exceeding the criterion is estimated by the false alarm rate. The criterion is assumed to be under the subject's control, and thus an important feature of sensory SDT experiments is explicit manipulation of the criterion via payoff matrices or presentation probabilities of signal and noise trials. A payoff matrix defines the costs for errors (false alarms and misses) and rewards for correct responses (hits and correct rejections).

SDT With Normal Distributions

In the classic version of SDT, both distributions are assumed to be unit normal curves. In an alternate version described in more detail later, the distributions are assumed to be logistic in form. The following indices are based on normal distributions. The corresponding indices for logistic distributions are given later.

Discrimination index (d'). According to this model, a subject's ability to discriminate between old and new items is given by d', the distance between the means of the old and new distributions in units of the common standard deviation. We assume throughout that the standard deviation of the old distribution, σ_o, equals the standard deviation of the new distribution, σ_n, and denote this common standard deviation as σ. This theoretical definition of d' can be expressed as

$$d' = (\mu_o - \mu_n)/\sigma,$$

where μ_o is the mean of the old distribution and μ_n is the mean of the new distribution.

Given a hit rate <1 and a false alarm rate >0, and the unit normal assumption, d' can be estimated by the z score of the false alarm rate minus the z score of the hit rate, or

$$d' = z_{FA} - z_H, \quad (1)$$

where z_H is the z score in the old distribution having H proportion above it and z_{FA} is the z score in the new distribution having FA proportion above it. For moderate biases ($H > .5$ and $FA < .5$), z_{FA} will be positive and z_H will be negative, so that Equation 1 has the effect of adding their absolute values.

Bias indices (β and C). This model generates several plausible bias measures, each with its own theoretical rationale. The purpose of any bias measure in this model is to locate the criterion, x_c, which is a dimensionless quantity. We consider two bias measures for the SDT model. The likelihood ratio measure, β, locates the criterion by the ratio of the heights of the old and new distributions, while the intersection measure, C, locates the criterion by its distance from the intersection of the two distributions, as shown in Figure 2.

The first bias measure, β, is the ratio of the density of the old distribution at the criterion divided by the same density for the new distribution. It is calculated as

$$\beta = f_o(z_H)/f_n(z_{FA}),$$

where f_o is the height of the normal distribution over z_H and f_n is the height of the normal distribution over z_{FA}.

This bias index is known as the likelihood ratio because it is the ratio of the likelihood of obtaining an observation equal to
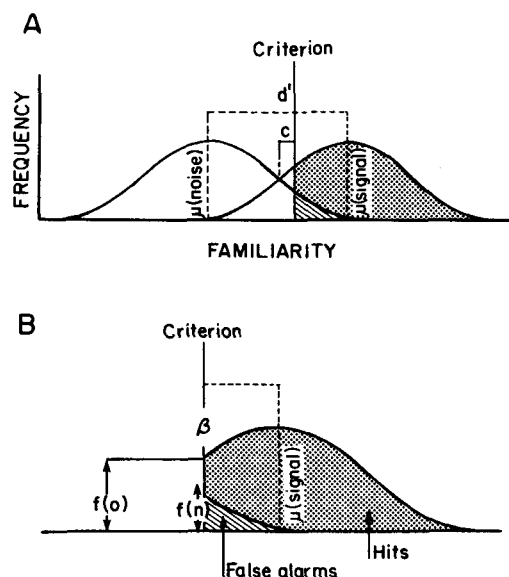

Figure 2. Definitions of C (A) and β (B) in the signal detection model.

the criterion given an old item to the likelihood of obtaining this observation given a new item. A subject with a perfectly neutral response bias will set his criterion at the intersection of the old and new distributions, yielding a β of 1. More liberal criteria lie to the left of the intersection point, yielding βs <1, whereas more conservative criteria lie to the right of the intersection point, yielding βs >1. Because symmetric placements of the criterion around the intersection points produce reciprocal values of β, β must be transformed before analysis to produce interval-scale data. The usual transformation is to take the natural logarithm, as shown by

$$\ln(\beta) = \ln[f_o(z_H)f_n(z_{FA})]. \quad (2)$$

Thus, ln β is 0 for a neutral criterion, negative for liberal criteria, and positive for conservative criteria.

Green and Swets (1974) showed that β is appropriate if a subject wishes to maximize expected winnings under a given payoff matrix. However, Lockhart and Murdock (1970) pointed out several problems with this measure when heterogenously memorable stimuli are used in an experiment. We will consider their reservations in more detail later.

An alternative way of locating x_c is to define its location relative to some zero point along the familiarity axis. One possible zero point is the mean of the new distribution. This locates the criterion within the new distribution, without regard to its location within the old distribution, and as such is determined only by the FA rate. Banks (1970) has termed this criterion index C_j and points out that it is only valid for hit/false alarm pairs which lie on the same isomemory function (that is, for points indicating equal discrimination).

The criterion index C_j can easily be calculated as z_{FA}, the z score having probability FA above it in a normal distribution. However, because we will be concerned with bias measured across different levels of discriminability, we do not consider this measure further, except to note that it has been used in the abnormal memory literature (e.g., Mohs & Davis, 1982).

A more useful measure of bias, which we call C, locates the criterion relative to the intersection of the old and new distributions. The intersection point defines the zero point, and distance from the criterion is measured in z score units. A completely neutral bias has a C value of 0, conservative biases produce positive C values, and liberal biases produce negative C values.

To compute C, we first determine the distance of the criterion from the mean of the new distribution, which is set equal to 0. This is given by z_{FA}, the z score having probability above it equal to the false alarm rate. To change the zero point from the mean of the new distribution to the intersection of the old and new distributions, we subtract $d'/2$ from z_{FA}:

$$C = z_{FA} - d'/2 = 0.5 (z_{FA} + z_H). \tag{3}$$

C and β compared. Lockhart and Murdock (1970) pointed out several problems with the use of β as a criterion measure. When heterogenously memorable stimuli (such as high- and low-imagery words) are presented for test, investigators usually assume that the criterion is set with respect to the appropriate old and new familiarity distributions; that is, the familiarity of a high-imagery word is compared to the distribution of both old *and* new high-imagery words. This assumption, however, requires that the subject first classify the item as a member of the appropriate class and then somehow locate its relevant new distribution before making such a decision. Yet locating the relevant new distribution would appear to require prior perfect knowledge of whether the item is old or new, the very decision we are asking the subject to make as the test of learning.

In contrast, the use of C requires only that the subject know the familiarity value of the test item. If we assume, with Glanzer and Adams (1985), that old and new distributions mirror one another in their locations along the familiarity axis, so that new distributions of highly memorable classes of items are lower on the familiarity continuum than new distributions of less memorable classes, this provides a mechanism whereby subjects can use the familiarity value itself as the basis for making the decision. If the increment in oldness is equal to the increment in newness, then the C index of criterion, defined as it is with respect to the intersection of the corresponding old–new distributions, appears to reflect a psychologically plausible mechanism.

SDT with Logistic Distributions

Noreen (1977) has shown that when logistic distributions replace normal distributions in a signal detection model, discrimination and bias indices are identical to those from two other important models of choice behavior: Luce's (1959, 1963) choice model and Link and Heath's (1975) random walk model. Additionally, indices based on SDT with logistic distributions are considerably easier to calculate than d', β, and C, thereby reducing the computational labor and chance for error. For these two reasons, we present these alternative discrimination and bias indices here. Later we show that the two sets of SDT-based indices yield equivalent results when applied to a variety of data sets.

The logistic distribution function (the cumulative of the density function) in standard form ($\mu = 0$, $\sigma = \pi/3^{0.5}$) is

$$F(x) = (1 + e^{-x})^{-1} = e^x/(e^x + 1),$$

where x is a standard score similar to the z score of the normal distribution.

The beauty of the logistic distribution is the simplicity with which the distribution function can be related to x (this is, of course, not a property shared by the normal distribution). By taking the logit, or log-odds transformation, of $F(x)$, we are left with x itself:

$$\ln [F(x)/(1 - F(x)] = \ln \{[e^x/(e^x + 1)]/[1/(e^x + 1)]\}$$
$$= \ln e^x = x.$$

The recovered score, x, is a standard score comparable to z for the normal distribution.

To recapitulate, in the logistic distribution, the (natural) log of the ratio of the area above an x score to the area below the x score produces the score itself. So, for example, the log of the hit rate divided by its complement the miss rate produces the value of x corresponding to the criterion, x_c, within the old item distribution. The corresponding operation on the false alarm rate and its complement (the correct rejection rate) produces the value of x_c within the new distribution. This property of the logistic distribution is used to derive the following measures (details of the derivations can be found in Noreen, 1977):

Discrimination index (logistic). A d'-like measure, d_L, is computed as follows:

$$d_L = \ln\{[H(1 - FA)]/[(1 - H) FA]\}. \tag{4}$$

Bias indices (logistic). A β-like measure can be computed as follows (the natural logarithm is computed for the same reasons as for the SDT normal model):

$$\ln(\beta_L) = \ln\{[H(1 - H)]/[FA(1 - FA)]\}. \tag{5}$$

A C-like measure can be computed as follows:

$$C_L = 0.5 [\ln\{[(1 - FA)(1 - H)]/[(H)(FA)]\}]. \tag{6}$$

Threshold Theories of Recognition Memory

Threshold models, in contrast to signal detection models, do not assume a continuum of memory strengths, but rather define discrete memory states. In one-high-threshold theory, one threshold defines two memory states: recognition and nonrecognition. If an old item crosses the subject's memory threshold, it will be correctly identified as old. If an old item fails to exceed the threshold, it may be identified as old or new depending upon the subject's response bias. New items can never cross the threshold in this model; this is why the model is termed "high" threshold. New items can only be misidentified as old by guessing on the basis of response bias from the nonrecognition state.

In a two-high-threshold model, there are two memory thresholds, one for old items and one for new items. Two thresholds define three possible memory states: old recognition, new recognition, and uncertainty. Old items crossing the old recognition threshold will always be identified as old and new items crossing the new item threshold will always be identified as new. As in the one-high-threshold model, new items can never cross the old item threshold and old items can never cross the new item threshold. Items in the uncertain state will be classified as

old or new depending on the subject's response bias. Thus, false alarms and misses always occur from the uncertain state. Because the one-high-threshold model is easily falsified by data, we consider here only the two-high-threshold model.

Two-High Threshold Model

Discrimination index (P_r). There are two discrimination indices in this model: P_o, the probability that an old item will exceed the old recognition threshold, and P_n, the probability that a new item will exceed the new recognition threshold. From a single hit and false alarm rate, it is impossible to obtain separate estimates for the two thresholds. Accordingly, we make the simplest possible operational assumption—that the two thresholds are equal—and define this common threshold as P_r.

The mirror effect in recognition, documented by Glanzer and Adams (1985), suggests that this equivalence assumption may be warranted. The mirror effect is the finding that as hit rates increase across various manipulations, the corresponding false alarm rates decrease in a "mirror" or inverse fashion.

The hit rate is composed of a certain proportion of true recognitions plus lucky guesses from the uncertain state. Because false alarms are only generated from the uncertain state, the false alarm rate is a direct estimate of the probability of saying "yes" when uncertain. Thus, the hit rate is related to P_r by

$$H = P_r + FA.$$

By rearranging terms, we get an estimate of P_r as

$$P_r = H - FA. \qquad (7)$$

Bias index (B_r). We define the bias index in this model, B_r, as the probability of saying "yes" to an item when in the uncertain state. False alarms occur when two things happen: The subject fails to recognize the item as new (which occurs with probability $1 - P_r$), and the subject guesses "yes" when uncertain (which occurs with probability B_r). Thus, the false alarm rate is related to B_r by

$$FA = (1 - P_r)B_r.$$

Rearranging terms and replacing P_r, the discrimination index, by $H - FA$, its estimate, gives us the following expression for B_r expressed in terms of the hit and false alarm rates:

$$B_r = FA/[1 - (H - FA)] \qquad (8)$$

A value of B_r equal to 0.5 indicates neutral bias, a value greater than 0.5 indicates liberal bias, and a value less than 0.5 indicates conservative bias.

It should be noted that the discrimination index from the two-high-threshold model is used quite frequently in recognition memory studies (e.g., Gillund & Shiffrin, 1984; Tulving & Thomson, 1971) although usually without reference to its underlying model. Indeed, as early as 1938, Woodworth suggested this particular "correction for guessing." Egan (1958) was apparently the first to state the model underlying P_r. In contrast, this is the first time, to our knowledge, that the bias measure B_r has been defined.

Distribution-free (Nonparametric) Model

Pollack and Norman (1964) proposed as an alternative measure of discrimination the area under an "average" isomemory
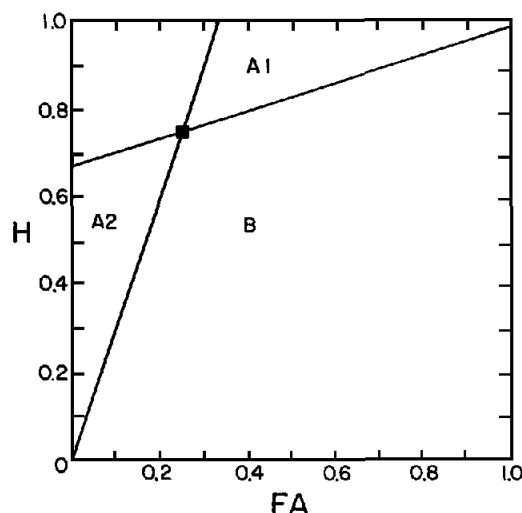

Figure 3. Areas in the unit square used to define the nonparametric indices.

curve drawn through a single pair of hit and false alarm rates. Use of the area under an isomemory curve follows from the demonstration by Green and Moses (1966) that the area under an isomemory curve is a good estimate of forced-choice memory performance. Figure 3 presents the areas used in deriving these nonparametric indices, and the computing formulas are from Grier (1971).

Discrimination index (A'). The rationale behind the computation of A' is that any reasonable isomemory function through the single hit/false alarm point shown in the unit square must pass through the areas $A1$ and $A2$. Accordingly, A' is defined as area B (all of which must lie below any reasonable isomemory function) plus 1/2 the sum of $A1$ and $A2$, or

$$A' = B + (A1 + A2)/2.$$

It is computed from a pair of hit/false alarm rates as follows: for $H \geq FA$,

$$A' = 0.5 + [(H - FA)(1 + H - FA)]/[(4H(1 - FA)]. \qquad (9A)$$

When $FA > H$, the point lies below the chance diagonal, and the formula must be revised as follows:
for $FA > H$,

$$A' = 0.5 - [(FA - H)(1 + FA - H)]/[4FA(1 - H)]. \qquad (9B)$$

Note that when $H = FA$, the point lies on the chance diagonal, and $A' = 0.5$.

Bias indices (B'' and B'_H). Several logical bias indices may be constructed. Grier (1971) proposed an index called B'', which is the difference in the two areas $A1$ and $A2$, divided by their sum. Notice that if $A1 < A2$, the subject has a bias toward "yes" and if $A2 < A1$, he has a bias toward "no." As for A', there are two computing formulas: one for the case of $H > FA$ and one for the case of $FA > H$. For $H \geq FA$,

$$B'' = [H(1 - H) - FA(1 - FA)]/[H(1 - H) + FA(1 - FA)]. \qquad (10A)$$

Table 1
Isomemory and Isobias Functions for Four Models of Recognition Memory

Model	Function
SDT with logistic distributions and β_L	
Isomemory	$H = aFA/[(a - 1)FA + 1]$, where $a = e^{d_L}$
Isobias for β_L	$H = \{1 + [1 + 4\beta_L(FA^2 - FA)]^{0.5}\}/2$
SDT with logistic distributions and C_L	
Isomemory	$H = aFA/[(a - 1)FA + 1]$, where $a = e^{d_L}$
Isobias for C_L	$H = (1 - FA)/(aFA - FA + 1)$, where $a = e^{2C_L}$
Two-high-threshold theory	
Isomemory	$H = P_r + FA$
Isobias	$H = [(B_r - 1)/B_r]FA + 1$
Distribution-free (nonparametric) theory with B''	
Isomemory	$H = \min\{1, [FA(1 - FA) + k^2]^{0.5} - k\}$, where $k = 1.5 - 2[FA + A'(1 - FA)]$
Isobias with B''	$H = 0.5 \pm \{0.25 - [FA(1 - FA)(1 + B'')/(1 - B'')]\}^{0.5}$

Note. SDT = signal detection theory.

When $FA > H$, the H and FA values are exchanged as follows:

$$B'' = [FA(1 - FA) - H(1 - H)]/[FA(1 - FA) + H(1 - H)]$$

(10B)

A second measure, proposed by Hodos (1970) is the difference between the two areas divided by A1, computed by

$$B'_H = 1 - \{[FA(1 - FA)]/[H(1 - H)]\}$$ (11A)

when $H \leq (1 - FA)$ so that bias is conservative and

$$B'_H = \{[H(1 - H)]/[FA(1 - FA)]\} - 1$$ (11B)

when $H > (1 - FA)$ so that bias is liberal.
When $H = FA$, $B'_H = 0$.

For both measures, a zero value indicates a neutral criterion, a positive value indicates a liberal criterion, and a negative value indicates a conservative criterion. Both bias measures lie between -1 and $+1$.

The computing formulas for these discrimination and bias measures are summarized in the Appendix.

Comparison of Models

Because three of the four basic models of recognition memory have two alternative bias measures, there are seven models defined by a discrimination-bias index pair. Each of these seven models defines a unique set of isomemory and isobias curves. Table 1 gives the formulas for the isomemory and isobias curves for four of the seven models, and Figure 4 shows isomemory and isobias functions for each model at two levels of discrimination and three levels of bias. The four models are SDT with logistic distributions and the bias measure β_L, SDT with logistic distributions and the bias measure C_L, two-high-threshold theory, and nonparametric theory with the bias measure B''. Unlike the SDT logistic model, the SDT normal model does not yield closed forms for the isomemory and isobias functions. Therefore, we have omitted SDT with normal distributions, which has the same shape as SDT with logistic distributions. We have also omitted nonparametric theory with B'_H, which has the same shape as that with B''.

The isomemory and isobias functions differ in shape, and therefore define different pairs of hit and false alarm rates which yield equivalent discrimination and equivalent bias. With the exception of two-high-threshold theory, which predicts linear isomemory functions, all other models predict curved isomemory functions. At high levels of discrimination, the nonparametric isomemory functions asymptote to the hit and false alarm axes quickly. This "flattening" means that there are minimum values of hit rates and maximum values of false alarm rates which limit the attainment of any given value of A'. For example, it is impossible to achieve an A' of .9 with a hit rate less than .6 or a false alarm rate greater than .4 (see Figure 4D).

The most notable differences among the models occur in the isobias contours. Only the isobias contours for two-high-threshold theory and for SDT with the intersection measure C maintain separation as discrimination decreases (that is, as the isobias contours approach the chance diagonal). Thus, with these models it is possible to observe bias differences among subjects even when they are operating close to chance. The other models (i.e., SDT with the likelihood ratio β and the nonparametric model with either bias index) predict isobias contours which converge toward common origins with chance performance. What this means in practice is that bias differences are harder to measure as performance decreases. In the limit, subjects performing at chance who show a maximum yea-saying bias (i.e., 100% hits and false alarms) or a maximum nay-saying bias (i.e., 0% hits and false alarms) will not be distinguishable from subjects showing intermediate levels of bias (including those showing no bias).

To further investigate the theoretical relation between discrimination and bias for each model, we varied discrimination but kept bias at its maximum possible values (both liberal and conservative). When discrimination is maximum (that is, 100% hits and 0% false alarms), there can be no bias regardless of the model. However, when discrimination is at chance (either 100% hits and false alarms, for a maximum liberal bias, or 0% hits and false alarms, for a maximum conservative bias) we should, in principle, be able to observe this with a reasonable bias index. Yet as is shown in Figure 5, two models (SDT with ln β and the nonparametric model with B'') show marked dependence of bias on discrimination at low levels of discrimination. That is, these maxima decrease as discrimination decreases. In the
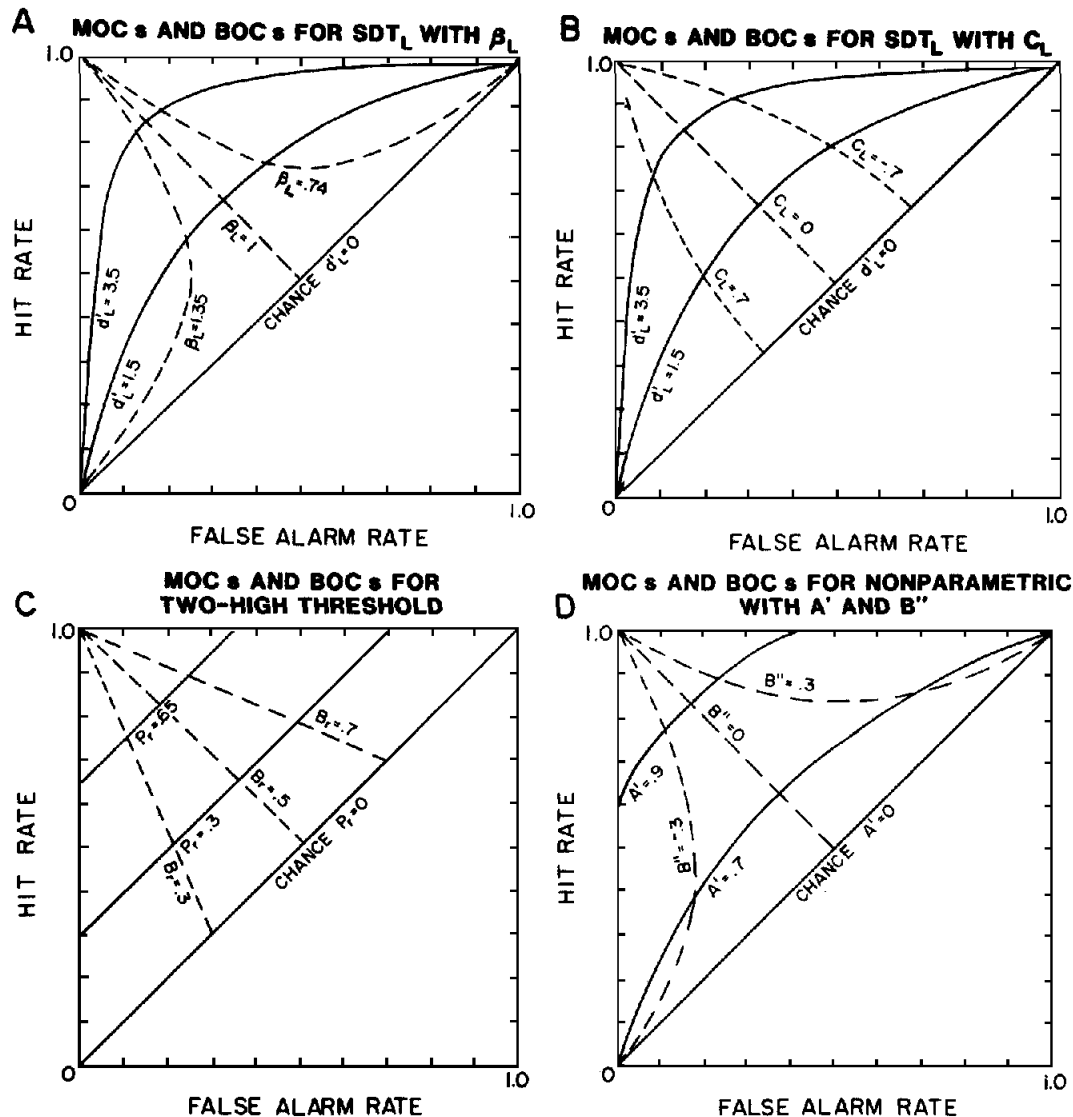
Figure 4. Isomemory and isobias functions for four models of recognition memory.
(MOC = memory operating characteristic; BOC = bias operating characteristic)

limit, as discrimination decreases to chance, each bias index approaches its neutral value. Thus, these recognition memory models show nonindependence between bias and discrimination in the sense that a decrease in discrimination decreases the range of values attainable by the bias measure.

On the basis of these results alone, we are tempted to recommend either SDT with the intersection measure C or the two-high-threshold model for use in recognition memory research, particularly in studies in which discrimination varies as a function of subject population or of experimental manipulation. It has sometimes been assumed that the distribution-free model is preferable on the grounds that it does not make assumptions about the form and nature of the process underlying generation of hits and false alarms. However, because it determines a unique isomemory and isobias function for any possible H, FA pair, we believe it constitutes as strong a model of recognition

memory as any other theoretical approach. Furthermore, as we have shown, it shows a marked dependence of bias on discrimination, and therefore violates our independence criterion.

In summary, at a theoretical level, three of the seven models—the two-high-threshold model and both SDT models with the intersection bias measure—fulfill our criterion of independence between discrimination and bias. In the first experiment, we address the pragmatics of whether theoretical lack of independence produces empirical problems. That is, over the range of levels of performance found in normal subjects across variations in stimulus materials, do the different models mirror their theoretical behavior in actual data?

Experiment 1

In Experiment 1, discrimination was manipulated in a recognition memory task by using high- and low-imagery word stim-
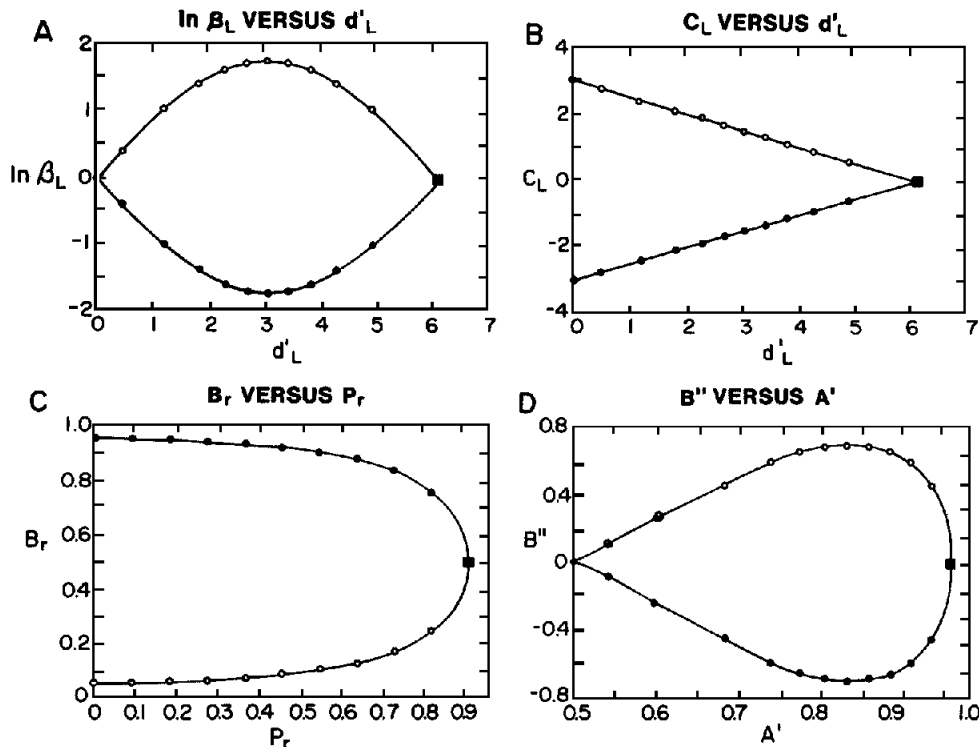
Figure 5. Relation between bias and discrimination for four models of recognition memory. (Open circles represent the maximum possible conservative bias; closed circles represent the maximum possible liberal bias.)

uli, and bias was manipulated by using three payoff matrices designed to produce liberal, neutral, and conservative response biases.

Method

Subjects and Design

A total of 10 subjects (7 male, 3 female) volunteered to participate in the experiment as part of an introductory psychology course requirement. The experiment had a 2 (imagery level: low and high) by 3 (payoff matrix: liberal, neutral, and conservative) factorial within-subjects design.

Materials

A total of 360 words were selected from the imagery norms published by Paivio, Yuille, and Madigan (1968) so as to be equally divided between high and low imagery levels. The high-imagery words were selected from those having scale values of 6 or above on a 7-point scale, and the low-imagery words were selected from those having scale values of 4 or below. All of the words were of A or AA frequency in the Thorndike and Lorge (1944) frequency counts.

The 360 words were divided into three sets of 120 each, of which half served as old items and half served as new items, with an equal number of high- and low-imagery items in the old and new sets. Words were typed in uppercase letters on 3 × 5 in. index cards.

Procedure

Each subject participated in three sessions on three successive days, each day under a given payoff matrix. The payoff matrices were pre-

sented to the subjects on index cards at the beginning of each test phase, and remained in view throughout. The neutral payoff matrix rewarded each hit and correct rejection by one point and penalized each false alarm and miss by one point. The liberal matrix rewarded hits more than correct rejections (+10 vs. + 1), and penalized misses more than false alarms (−10 vs. − 1). The conservative matrix did just the opposite, rewarding correct rejections more than hits (+10 vs. +1) and penalizing false alarms more than misses (−10 vs. −1). Subjects were motivated to earn a maximum total number of points by the offer of a monetary prize for the best performance.

Each subject participated in three sessions on three successive days, one day under each of the payoff matrices. The test procedure was identical on each day. Subjects were first presented with one set of 60 study items (30 low and 30 high imagery) and given 2 min to go through the pack at their own pace. This study period was followed by a 2-min filled delay. Subsequently, subjects were presented with the day's payoff matrix, which was explained by the experimenter. The subject was then presented with a shuffled deck of old and new items and asked to sort the items into appropriate piles. The test phase was untimed. At the end of each day's session, test performance was calculated and reported to the subject. The orders of word sets and payoff matrices were counterbalanced across subjects.

Results and Discussion

Figure 6 presents hit/false alarm rate pairs for each of the six conditions pooled across subjects. The imagery and payoff manipulations had the intended effects: High-imagery words were discriminated better than low-imagery words, and the liberal payoff matrix produced more liberal responding and the
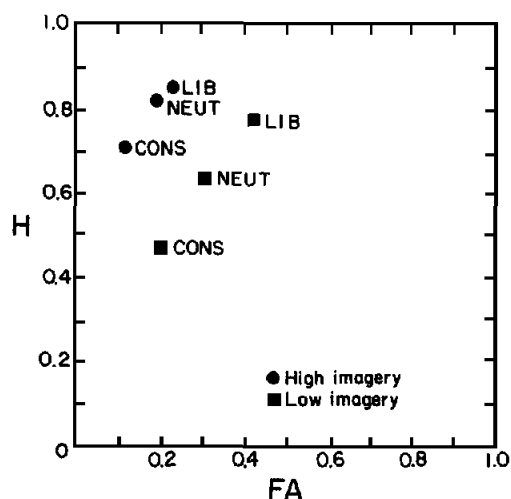
Figure 6. Hit and false alarm rates for Experiment 1.

conservative payoff matrix more conservative responding than the neutral matrix. The empirical isobias functions appear to converge to the chance diagonal, rather than to the origins, thus supporting either the SDT model with the C intersection bias parameter or the two-high-threshold model. We present statistical evidence for this observation later.

Comparison of Models

For each model, all possible pairs of discrimination and bias indices were calculated and submitted to 2×3 within-subjects analyses of variance (ANOVAs) with imagery and payoff matrix as the two independent variables and index as the dependent variable. An ideal pair of indices would show the following properties: For the discrimination index, (a) a significant effect of imagery, (b) no effect of payoff, and (c) no interaction; for the bias index, (a) a significant effect of payoff, (b) no effect of imagery, and (c) no interaction.

Tables 2 and 3 present mean values of discrimination and bias indices for each of the models. The measures are grouped by the variable across which they are expected to remain constant. Thus, discrimination measures for the high-imagery condition for all three payoff matrices are to the left in Table 2, and discrimination measures for the low-imagery condition for all three payoff matrices are to the right. Superscripts to each measure indicate the degree to which it fulfills the *three* conditions.

Considering first the discrimination indices (Table 2), neither hits nor false alarms alone satisfy all three conditions, but all other discrimination indices do. Thus, behavior of the discrimination measures does not permit us to reject any of the "reasonable" models (neither hits alone nor false alarms alone correspond to discrimination measures for any reasonable model).

The bias indices in Table 3 are grouped by payoff matrix. The two bias measures to the left are for the conservative matrix, those in the middle are for the neutral matrix, and those on the right are for the liberal matrix. Here several of the bias indices fail the independence test by showing a significant interaction. In particular, $\ln \beta$ from both SDT models fails, as do both distri-

bution-free measures. The nature of this interaction is just what we would expect on the basis of the theoretical relations: namely, that bias is more extreme for high-imagery words than for low-imagery words even though subjects made their yes–no decisions to random presentations of the two types of words.

The two models that pass the independence test are two high threshold and SDT (normal and logistic) using the interesection bias index C. Thus on the basis of these results we would recommend the use of either two-high-threshold or SDT measures with the stipulation that C be used in place of $\ln \beta$. This deviates from common practice because most investigators measure response bias with β. These empirical results confirm our previous observations on the theoretical nonindependence of d' and $\ln \beta$ and of A' and either B'' or B'_H.

Experiment 2

In the second experiment, we applied the preferred measures of discrimination and bias to yes–no recognition memory for meaningful pictures in young normal and moderately to moderately severely demented subjects. Both groups received repeated study–test trials with a final delayed test trial. We examined the fate of the models in this paradigm, and explored the differences between normal and abnormal picture recognition memory.

In addition to our previous concerns, we were interested in recognition paradigms for practical reasons. The recent intensive efforts to pharmacologically ameliorate age-related cognitive decline (viz., Bartus, Dean, Beer, & Lippa, 1982) suffer from a lack of appropriate tasks for more severely impaired subjects (Semple, Smith, & Swash, 1982). Additionally, Brinkman and Gershon (1983), in a review of methods for measuring cholinergic drug effects on memory, suggested that visual recognition tasks may be particularly useful in this regard.

Method

Subjects and Design

The young normals were 101 undergraduates who volunteered to participate in the experiment as part of an introductory psychology

Table 2

Mean Values of Discrimination Indices for the Four Basic Models in Experiment 1

	High imagery			Low imagery		
Model	C	N	L	C	N	L
Hit rate[a,c]	.716	.826	.868	.474	.632	.723
False alarm rate[a,c]	.097	.161	.219	.203	.303	.397
d'(SDT:norm)[a,b,c]	2.013	2.090	2.148	0.818	0.897	0.935
d'(SDT:log)[a,b,c]	3.480	3.580	3.738	1.381	1.467	1.553
P_r(2HT)[a,b,c]	.619	.664	.648	.271	.329	.326
A'(nonpar)[a,b,c]	.880	.895	.886	.708	.731	.724

Note. C = conservative payoff; N = neutral payoff; L = liberal payoff; SDT = signal detection theory; 2HT = two-high-threshold theory; nonpar = nonparametric theory.
[a] Significant effect of imagery ($p < .05$). [b] No significant effect of payoff. [c] No interaction.

Table 3
Mean Values of Bias Indices for the Four Basic Models in Experiment 1

	Payoff					
	Con-servative		Neutral		Liberal	
Model	H	L	H	L	H	L
Hit rate[a,c]	.716	.474	.826	.632	.868	.723
False alarm rate[a,c]	.097	.203	.161	.303	.219	.397
ln β (SDT:norm)[a,b]	.635	.350	.077	.016	−.265	−.061
C(SDT:norm)[a,b,c]	.363	.485	.032	.085	−.161	−.162
ln β_L(SDT:log)[a,b]	.777	.435	.093	.020	−.324	−.077
C_L(SDT:log)[a,b,c]	.659	.816	.063	.137	−.290	−.265
B_r(2HT)[a,b,c]	.275	.281	.475	.464	.597	.572
B'' (nonpar)[a,b]	.354	.224	.043	.019	−.148	−.041
B_H(nonpar)[a,b]	.490	.285	.060	.015	−.210	−.073

Note. H = high imagery; L = low imagery; SDT = signal detection theory; 2HT = two-high-threshold theory; nonpar = nonparametric theory.
[a] Significant effect of payoff ($p < .05$). [b] No significant effect of imagery. [c] No interaction.

course requirement. Approximately 20 subjects were tested with each form of the memory test.

The demented subjects were 11 patients (10 male, 1 female) at the New York Veterans Administration Medical Center diagnosed as having either Alzheimer's disease ($n = 9$) or parkinsonian dementia ($n = 2$). Diagnosis was made according to the *Diagnostic and Statistical Manual of Mental Disorders III* (American Psychiatric Association, 1980) and National Institute of Neurological and Communicative Disorders and Stroke—Alzheimer's Disease and Related Disorders Association (McKhann, Drachman, Folstein, & Katzman, 1984) criteria for primary degenerative dementia or dementia with Parkinson's disease. All subjects had moderate to moderately severe memory and cognitive dysfunction with Global Deterioration Scores (GDS) of 3 (early confusional stage) to 6 (middle dementia stage) as evaluated by the Guild Memory Test (Crook, Gilbert, & Ferris, 1980) and the Brief Cognitive Rating Scale (Reisberg, Ferris, & Crook, 1982). Of the 11 subjects, one was classified as GDS 3, three as GDS 4, five as GDS 5 and two as GDS 6. Subjects ranged in age from 55 to 79. No subject was receiving treatment with concomitant anticholinergic medication.

Subjects were tested as part of approved treatment protocols for novel therapeutic agents for dementia. None of these investigational new drugs yielded positive therapeutic effects. Subjects were tested repeatedly on from two to five different equivalent forms of the picture memory test, for a total of 119 administrations.

The design for the young normals was a 5 (form) by 3 (test trial) mixed design with form between subjects and test trial within subjects. The design for the demented patients was a 5 (form) by 4 (test trial) mixed design with the same designation. Thus, demented patients received an additional study–test trial. In addition, the data from the demented and normal subjects were analyzed together (by omitting Trial 3 for the demented subjects and collapsing across form) to produce a 2 (subject group) by 3 (test trial) mixed design.

Materials

The stimuli consisted of 250 pictures selected from the Snodgrass and Vanderwart (1980) set of 260 meaningful picture stimuli. These stimuli consist of black-on-white line drawings of real objects (e.g., asparagus,

sailboat). Five forms of the test were constructed for the patient subjects. Each form contained one set of 10 target items, and four sets of 10 distractor items. Items were chosen such that each form and set of stimuli were balanced for familiarity, visual complexity, and category membership according to the Snodgrass and Vanderwart (1980) and Battig and Montague (1969) norms. To construct the set used with normal subjects, the 50 items from each form for the patients were concatenated to produce five sets of 50 target ("old") items. The 50 target items were paired with three sets of 50 distractor ("new") items so that each normal subject saw 200 of the 250 stimuli.

Procedure

All subjects were tested in a repeated study–test procedure followed by a delayed test trial. The procedure for the patients was modified for the normal subjects to avoid ceiling effects for the latter group. A study phase consisted of presenting the set of n study stimuli (where $n = 10$ for patients and 50 for normals) presented via cards for 5 s each accompanied by the picture's name (patients) or presented via slide projector for 1 s each with no name (normals).

The test phase consisted of presenting the n study items as targets interspersed with an equal number of distractors and asking subjects, for each item, to indicate whether the item was old (by responding "yes") or new (by responding "no"). Test stimuli were presented on cards (for patients) or slide projector (for normals) and responses were oral (for patients) or written (for normals). For both groups, the order of the target stimuli in the test sequence was randomly different from the order on Study Trial 1, subject to the constraint that no more than three old or new stimuli could appear successively.

A second study–test sequence followed immediately for both groups of subjects. For the patients, the order of the targets remained the same as for Study Trial 1; for the normals, the order of the targets was rerandomized. Patients had a third study–test trial, but normals did not. Distractors for each test trial were always new for both groups.

Both groups were then given a filled 30-min delay which for the patients consisted of performing other memory tasks, and for the normals consisted of completing a standard test of spatial relations. At the end of the filled delay, both groups were given a surprise recognition test, in which the n targets were presented again mixed with an equal number of new distractors. At the end of the recognition test, normal subjects were additionally asked to recall as many of the targets as they could. These recall data will not be presented here.

Results and Discussion

Effect of Form

To analyze the effect of form for the patients, only those subjects receiving all five forms were used (7 of the 11), so as to avoid effects of heterogeneity of performance. A 4 (trials) by 5 (forms) within-subjects ANOVA was performed separately on the hit and false alarm rates. Neither form nor the Form × Trials interaction was significant for either analysis. For hits, the main effect of form yielded $F(4, 24) = 2.35$ and for the Form × Trial interaction, $F(12, 72) = 1.21$. For false alarms, the main effect of form yielded $F < 1$, and the Form × Trial interaction yielded $F(12, 72) = 1.76$. Therefore, all subsequent analyses for patient subjects will ignore form as a variable.

Each normal subject received only one form, so a 4 (trials) by 5 (forms) mixed ANOVA was performed separately on the hit and false alarm rates. Neither measure showed a significant main effect of form or a significant Form × Trial interaction. For hits, the main effect of form yielded an $F(4, 96) = 1.15$, and

Table 4
Mean Values of Discrimination Indices for Normal Subjects by Test Trial: Experiment 2

Index	Trial 1	Trial 2	Delay
Hit rate	.787	.952	.960
False alarm rate	.125	.041	.042
d'(SDT:norm)	2.09	3.70	3.76
P_r(2HT)	.662	.910	.917

Note. Both discrimination measures show significant changes toward better discrimination between Trials 1 and 2, but no change between Trial 2 and the delay trial. SDT = signal detection theory; 2HT = two-high-threshold theory.

for the interaction of form with trials, $F(8, 192) = 1.66$. For false alarms, the main effect of form had an $F(4, 96) = 1.80$, and for the interaction of form with trials, $F < 1$. Therefore, all subsequent analyses for normal subjects will ignore form as a variable.

Learning and Forgetting Across Trials

Recognition performance in terms of preferred discrimination and bias indices will first be analyzed for each group separately because of the different numbers of learning trials in the two groups, and then the two groups will be combined. The preferred indices are d' and C from the SDT model with normal distributions and P_r and B_r from the two-high-threshold model.

Normals' performance. Table 4 shows mean values of discrimination indices for the three trials, and Table 5 presents the same results for the bias indices.

Both d' and P_r demonstrate a highly significant effect of trials by ANOVA, $Fs(2, 200) > 237$, both $ps < .001$. Planned comparisons show that discrimination increases for both indices between Trials 1 and 2 and fail to decrease between Trial 2 and the delay trial.

Both bias indices also change significantly across trials, $Fs(2, 200) > 9.34$, both $ps < .0001$. Inspection of Table 5 suggests that normal subjects demonstrate conservative response biases on Trial 1, which become close to neutral on Trial 2 and slightly liberal by the delay trial. Planned comparisons show that this change from conservative to neutral is significant for both bias

indices between Trials 1 and 2, but not between Trial 2 and the delay trial for either bias index.

Patients' performance. Tables 6 and 7 present the same preferred discrimination and bias indices for patients as a function of trial. In the one-way repeated-measures ANOVAs performed, a conservative approach was taken, in which replications within subjects were collapsed to yield a single estimate for each of the 11 demented patients. This procedure yields a sample size representative of those used in studies with impaired populations.

Both d' and P_r show significant effects of trial, $F(3, 30) = 7.23$, and $F(3, 30) = 5.87$, respectively, both $ps < .01$. Planned comparisons were performed between Trials 1 and 2, 2 and 3, and 3 and the delay, with identical results for each dependent measure. For both, there were significant increases between Trials 1 and 2, but not between Trial 2 and Trial 3. Also, neither measure showed forgetting across the delay.

Of the two bias measures, only B_r showed a significant increase in liberality across trials, $F(3, 30) = 4.37, p = .012$. No pair-wise sequential planned comparisons on B_r were significant, and the decrease in C across trials failed to reach significance, $F(3, 30) = 2.34, p = .09$. Thus, in contrast to the discrimination measures, only B_r is sensitive to trials in showing that the apparent change toward more liberal bias in the patients is reliable.

Comparison of Patient and Normal Performance

In order to compare patients with normal subjects, we deleted Trial 3 from the patient data set and performed 2 × 3

Table 5
Mean Values of Bias Indices for Normal Subjects by Test Trial: Experiment 2

Index	Trial 1	Trial 2	Delay
Hit rate	.787	.952	.960
False alarm rate	.125	.041	.042
C(SDT:norm)	.180	.024	−.019
B_r(2HT)	.382	.487	.518

Note. Both bias indices show significant changes from conservative to neutral bias from Trial 1 to Trial 2, but no change between Trial 2 and the delay trial. SDT = signal detection theory; 2HT = two-high-threshold theory.

Table 6
Mean Values of Discrimination Indices for Patients by Test Trial: Experiment 2

Index	Trial 1	Trial 2	Trial 3	Delay
Hit rate	.782	.843	.897	.888
False alarm rate	.359	.326	.327	.330
d'(SDT:norm)	1.46	1.80	2.06	1.95
P_r(2HT)	.423	.517	.570	.558

Note. Both discrimination measures show significant changes towards better discrimination between Trials 1 and 2, but not between Trials 2 and 3 or Trial 3 and the delay. SDT = signal detection theory; 2HT = two-high-threshold theory.

Table 7
Mean Values of Bias Indices for Patients by Test Trial: Experiment 2

Index	Trial 1	Trial 2	Trial 3	Delay
Hit rate	.782	.843	.897	.888
False alarm rate	.359	.326	.327	.330
C(SDT:norm)	−.221	−.256	−.372	−.388
B_r(2HT)	.580	.599	.651	.685

Note. Only B_r showed a significant increase with trials; however, no pairwise comparisons were significant. SDT = signal detection theory; 2HT = two-high-threshold theory.
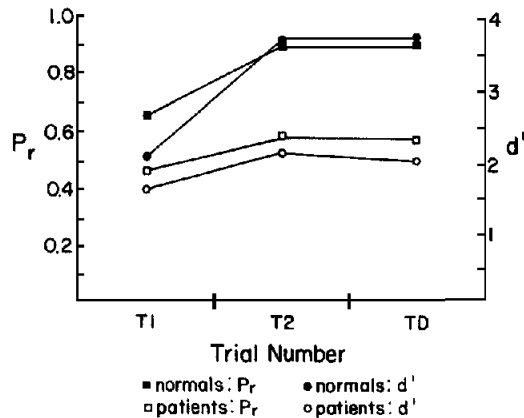
Figure 7. Discrimination values (P_r and d') for normal and demented subjects by trials for Experiment 2.

mixed-design ANOVAs on the preferred discrimination and bias indices, with the between-subjects variable diagnosis (demented vs. normal) and the within-subjects variable trial. To increase the power of the comparison, we selected the first administration of each form to each patient, and treated it as a separate subject in the analyses. This procedure generates 47 demented "subjects," compared with 101 normal subjects. Figures 7 and 8 show values of discrimination and bias indices for both groups of subjects as a function of trials.

As is apparent in Figure 7, normals performed better than patients on both discrimination measures, and there was a significant effect of trials, and a significant interaction between trials and diagnosis because of the faster learning of the normals (all $ps < .001$).

More interesting are the bias differences between the two subject groups shown in Figure 8. For both B_r and C, there was a main effect of diagnostic group, with the demented subjects showing much more liberal biases than the normals: for B_r, $F(1, 146) = 27.86$, and for C, $F(1, 146) = 45.44$, both $ps < .0001$. Additionally, the main effect for trials was significant for both bias measures: for B_r, $F(2, 292) = 22.81$ and for C, $F(2, 292) = 23.35$, both $ps < .0001$. The interactions between diagnosis and trial were not significant for either measure (both $Fs < 1$). Furthermore, although the normal subjects showed slightly conservative biases on average, the demented patients showed markedly liberal biases.

It is clear that we did not succeed in equating discrimination between demented and normal subjects by increasing the numbers of items in the old and new sets. Not only were normals superior to demented subjects on the first trial, but also they showed more rapid learning. Of more interest is the difference in the behaviors of the bias indices, with the demented subjects showing a robustly liberal bias, and the normals demonstrating an overall slightly conservative bias. Both groups become more liberal over trials in a parallel fashion.

Experiment 3

As Experiment 3, we report data from a study conducted by Butters et al. (1985) who used a yes–no recognition memory

task similar to that reported in Experiment 2, but with words instead of pictures. They compared the memory performance of patients with Huntington's disease and amnesia with that of elderly normal controls. (We wish to thank them for providing us with their raw data for this analysis.)

We examined these data to extend the work reported in Experiment 2 to see if the differences we see between normal and abnormal recognition memory in Alzheimer's disease and parkinsonian dementia obtain in Huntington's disease (another dementing illness) with different stimulus material, and to extend our analysis of discrimination and bias to amnesic patients.

Method

Complete descriptions of the subjects, materials, and procedure are given in Butters et al. (1985) and are briefly reviewed here.

Subjects and Design

Subjects were 9 amnesics of various etiologies (6 Korsakoff's psychosis amnesics, 2 postencephalitic patients, and 1 patient with a medial-temporal neoplasm), 10 patients with Huntington's disease with a range of functional disability, and 14 age-matched normal controls. The amnesics were older than the Huntington's disease patients. The experiment followed a 3 (diagnosis: amnesic, demented, and normal) by 6 (test trial) mixed design.

Materials

Two forms of a 15 target–15 distractor yes–no recognition task were constructed based on the Rey Auditory Verbal Learning Test (Lezak, 1983). Additional items to serve as distractors were chosen to match the target items in frequency.

Procedure

The procedure used in this study was identical to that of Experiment 2 with the following exceptions. Five immediate study–test trials were administered; a 10-s rest period was interposed between a test trial and the subsequent study trial; a 20-min filled delay intervened between the fifth test trial and the sixth (delayed) test trial.
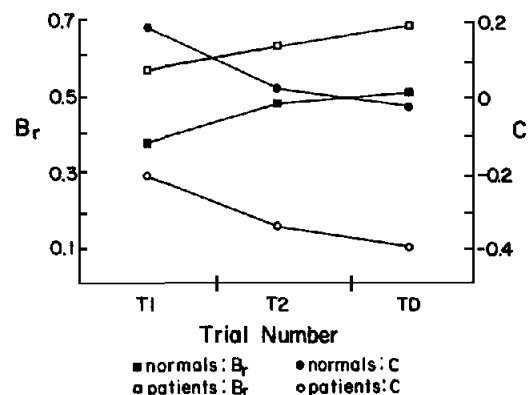

Figure 8. Bias values (B_r and C) for normal and demented subjects by trials for Experiment 2.

Table 8

Mean Values of Discrimination Indices for the Three Groups of Subjects in Butters, Wolfe, Martone, Granholm, and Cermak (1985)

Index	Normals	Amnesics	Huntington's disease
Hit rate	.922	.724	.867
False alarm rate	.043	.227	.210
d'(SDT:norm)	3.35	1.58	2.26
P_r(2HT)	.879	.497	.656

Note. Both discrimination measures show significant differences among groups. SDT = signal detection theory; 2HT = two-high-threshold theory.

Results and Discussion

Preferred discrimination and bias indices were computed as in Experiment 2. Discrimination indices averaged across trials for the three diagnostic groups are presented in Table 8; Table 9 gives the corresponding bias indices. These data were analyzed by mixed-design 3 (diagnostic groups) × 6 (trials) ANOVAs.

Discrimination is best for normal subjects, worst for amnesics and intermediate for the Huntington's disease patients. Both discrimination indices showed a significant main effect of subject group: for P_r, $F(2, 30) = 50.74$, and for d', $F(2, 30) = 71.56$, both $ps < .0001$. All of the planned pair-wise comparisons were significant, showing that normals were significantly better than Huntington's disease patients, who in turn were significantly better than amnesics.

Both measures also showed a highly significant effect of trials: for P_r, $F(5, 150) = 52.43$, and for d', $F(5, 150) = 50.43$, both $ps < .0001$. The interaction between diagnosis and trials was significant for P_r, $F(10, 150) = 3.64$, $p < .001$, but not for d', $F(10, 150) = 1.57$, ns. This interaction, shown in Figure 9, appears to be due to differential rates of forgetting between Trial 5 and the delay trial among the groups: the amnesics forgot the most, the Huntington's disease patients forgot a little, and the normals forgot not at all. Although d' also shows the same pattern, the interaction for d' failed to reach significance.

As is shown in Figure 10, the bias measures show a completely different pattern of results. The normal and amnesic subjects are somewhat conservative and do not appear to differ from one another, whereas the Huntington's disease patients show much more liberal bias than the other two groups.

Table 9

Mean Values of Bias Indices for the Three Groups of Subjects in Butters, Wolfe, Martone, Granholm, and Cermak (1985)

Index	Normals	Amnesics	Huntington's disease
Hit rate	.922	.724	.867
False alarm rate	.043	.227	.210
C(SDT:norm)	.096	.076	−.169
B_r(2HT)	.431	.455	.594

Note. Of the two bias indices, only B_r showed a significant effect of groups, with C only marginally significant. SDT = signal detection theory; 2HT = two-high-threshold theory.
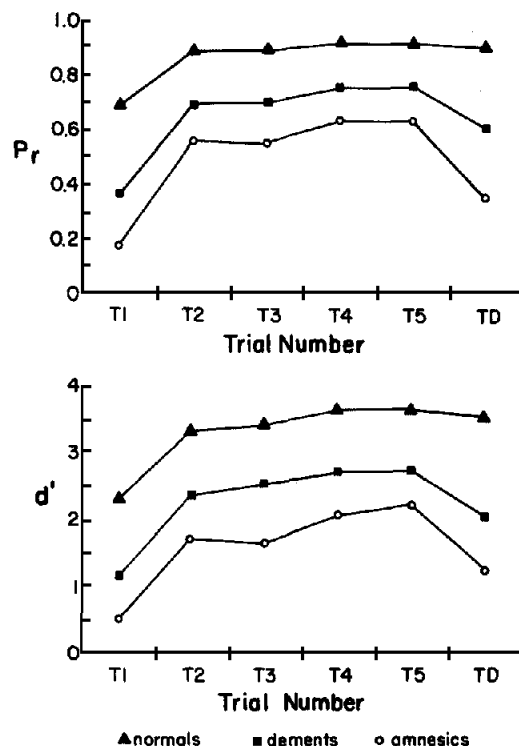

Figure 9. Discrimination indices (P_r and d') by trial for Butters, Wolfe, Martone, Granholm, and Cermak's (1985) subjects.

For B_r, the effect of diagnostic group was significant, $F(2, 30) = 3.73$, $p = .036$, with normals and amnesics showing somewhat conservative bias, and Huntington's disease patients showing liberal bias. There was a highly significant effect of trials, with bias becoming more liberal across trials, $F(5, 150) = 9.40$, $p < .0001$. The interaction was not significant ($F < 1$). For C, the differences among diagnostic groups failed to reach significance, $F(2, 30) = 3.00$, $p = .065$. There was, however, a highly significant effect of trials, $F(5, 150) = 9.45$, $p < .0001$, and no interaction ($F < 1$).

To recapitulate, both models show the following important results: In terms of discrimination, amnesic memory is worse than demented memory which in turn is worse than normal memory. The effect of trials was significant for both discrimination and bias. All groups learned (that is, showed increases in their discrimination indices), and all groups became more liberal over trials. More interestingly, demented bias was more liberal than amnesic or normal bias, which were equal. This last finding replicates our own results from Experiment 2 and a number of similar observations in the literature. We take these particular bias results quite seriously, because they seem to indicate that liberal bias is not solely a result of poor memory (or the amnesic patients would surely show it too). Thus, there is some property of dementia which causes new items to appear familiar, or old.

Turning next to the more subtle differences in patterns of results for the two preferred models, we found that although both discrimination indices showed a significant effect of group and of trials, only P_r showed a significant interaction, indicating that
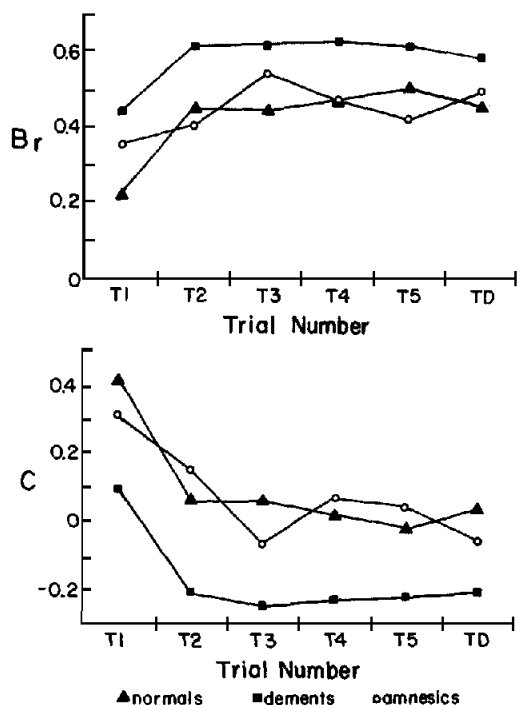
Figure 10. Bias indices (B_r and C) by trial for Butters, Wolfe, Martone, Granholm, and Cermak's (1985) subjects.

normal and demented subjects learn at a greater rate and forget less than the amnesics.

In addition, although bias measures showed a significant increase in liberality with trials, only B_r showed a significant effect of subject group, with demented subjects showing more liberal biases than either normal or amnesic subjects.

General Discussion

There are several major findings from these explorations of recognition memory. First, not all commonly used theories of the recognition memory process fare equally well in yielding independence of discrimination from response bias. The two models that pass the independence test are two-high-threshold and SDT models with the intersection bias index C rather than β, the ratio bias index.

Second, the models appear differentially sensitive to changes in bias and discrimination. Two-high-threshold measures are more sensitive to change than are SDT measures.

Third, all subjects, normal, demented and amnesic, demonstrated parallel increases in liberality of response bias across learning trials.

Finally, normal response bias distinguishes amnesia from the three types of dementia tested. Alzheimer's, parkinsonian, and Huntington's dementia were all marked by abnormally liberal bias in addition to poor discrimination.

The following discussion addresses the implications of each of the major findings in turn.

Acceptability of Models

Only two models of recognition memory yield indices of discrimination that are independent of bias. The acceptable theo-

ries are two-high-threshold with P_r and B_r, and SDT with normal and logistic distributions using d' and C.

When other indices are reported, experimenters are faced with such tasks as disambiguating changes in the bias index which appear as functions of changes in discrimination (e.g.: as with d' and β), or explaining changes in a bias index that is invalid under prevailing experimental conditions (e.g. as with C_j in repeated study–test paradigms) (for examples see Hart, Smith, & Swash, 1985; Mohs & Davis, 1982). Discussion of the meaning of results is simplified by use of either of the two models in which the experimenter is assured that memory and response strategy measures are independent.

Differential Sensitivity of Models

Discrimination

Although in Experiments 1 and 2, the discrimination measures from both models showed identical patterns of statistical significance, in Experiment 3, the two-high-threshold discrimination measure appears to be more sensitive than the SDT discrimination measure. Specifically, P_r showed a significant interaction between diagnostic group and trials as a result of the differential forgetting of the three groups between Trial 5 and the delay trial, whereas the same interaction failed to reach significance for d'.

Bias

A similar pattern of sensitivity holds for the bias measures from the two models. In Experiment 2, only the two-high-threshold bias measure, B_r, showed a significant increase in liberality across trials for the Alzheimer's disease and parkinsonian dementia patients, whereas this effect failed to reach significance with the SDT measure C. In Experiment 3, the differences among diagnostic groups in bias reached significance for B_r, but again, not for C.

What should the researcher do? As before, we believe the answer is a pragmatic one. We recommend the use of both acceptable models to characterize recognition memory under any manipulation. A negative result under one theory may be positive under another; both should be reported and discussed. If multiple nonindependent comparisons are problematic, stricter criteria for excluding chance significance may be used. Furthermore, if the experimental manipulation is designed to improve performance to a clinically relevant degree (as in trials of novel therapeutic agents for Alzheimer's disease), efficacy is best supported if improvement in seen in both two-high-threshold and SDT measures.

Increasing Liberality Over Trials

All subjects in Experiment 2, regardless of diagnosis, demonstrated increased liberality of response bias across learning trials in the absence of explicit manipulations of payoff matrix. This effect was most prominent between the first and second learning trials. Mohs and Davis's (1982) Alzheimer's disease subjects also showed this effect in C and B_r as we calculated them from the presented raw data. The simplest explanation of this widespread increase in liberality across learning trials is a

buildup of interference across repeated study–test trials. That is, as more items are presented, the net familiarity of all items shifts upward (in SDT terms), while subjects' absolute level of familiarity used as the criterion value remains fixed. This leads to an overall increase in both hits and false alarms without an actual downward shift of criterion. The liberality effect is more difficult to explain in two-high-threshold terms. In this case, the theory holds that increases in hits and false alarms are a result of increased guesses of yes when uncertain. That is, there must be a real change in response bias, rather than a simple upward shift in net familiarity. Thus, the SDT model is a more appealing conceptualization.

Response Bias and Diagnosis

Why might amnesic subjects have normal response bias in the face of the worst discrimination performance of all patients tested? That is, why do the demented subjects show the most liberal response bias when their discrimination is intermediate between normal and amnesic subjects?

Two major explanatory systems can be brought to bear on this question. A neurochemical explanation would focus on the neurotransmitter deficits characteristic of each disorder, and conclude that these differences sufficiently accounted for the different patterns of memory deficit. Cognitive psychological explanations would focus on the psychological processes which differed between groups. Ideally, the two types of explanation would coincide such that the learning process differences would be mediated by the neurotransmitter systems damaged in each disorder.

We turn first to a cognitive psychological explanation of the recognition memory process. A feature match model of recognition holds that recognition takes place via pattern matching between features of the test stimuli and stored active representations of the "old" items. If a match or near match is found, the subject responds "yes"; otherwise the subject responds "no." In terms of familiarity, an increasing number of features in common between the test items and internal representations (either from the target set or from semantic memory) yields a parallel increase in subjective familiarity. Thus, subjects might have high levels of familiarity for both old and new items if stimuli are not encoded distinctively at presentation and semantic memory representations have been inappropriately activated by prior target and distractor items.

Thus, stimuli that are not encoded distinctively at presentation will adequately match stored representations of similiarly poorly encoded target items or very familiar items in semantic memory, leading to a general increase in yes responses to both old and new items.

There is some evidence for this "distinctiveness" hypothesis of failure to distinguish between old and new items. It has been found repeatedly that Alzheimer's disease patients perform poorly on tasks such as category retrieval, object naming and similarities (e.g., Bayles, 1982; Ober, Dronkers, Koss, Delis, & Friedland, 1986). The most common interpretations of these results is that Alzheimer's disease produces disorders in semantic memory structure or access (Martin & Fedio, 1983; Nebes, Boller, & Holland, 1986; Ober et al., 1986; Warrington, 1975).

Two sets of findings are particularly germane. The most di-

rect demonstration of failure to encode distinctive features has been provided by Grober, Buschke, Kawas, and Fuld (1985). These investigators asked mildly impaired Alzheimer's disease subjects and matched normal controls to rate features of concepts, and found that the Alzheimer's disease patients did not distinguish as well between prototypical features (e.g., legs–chair) and non-critical features (e.g., cloth–chair). Additionally, Nebes et al. (1986) have shown that Alzheimer's disease patients are more sensitive to context effects on a sentence completion task in that these patients performed quite poorly unless the sentence frame highly constrained the set of potential responses. These results may be interpreted as a consequence of failure to locate appropriate features of potential responses given less than optimal search cues.

On the other hand, normal semantic memory and normal intelligence are defining features of amnesia (Moskovitch, 1982). Therefore, it is not surprising that priming, category retrieval, and similarities tasks are performed near normally by amnesics (Parkin, 1984).

Thus, whether at encoding, storage or retrieval, Alzheimer's disease patients show evidence of blurring of distinctive features of memoranda; to our knowledge, this dysfunction has not been reported in amnesia.

Neurochemically, a case has been made for the importance of cortical acetylcholine deficits in the production of abnormally liberal bias in recognition and intrusion errors in recall tasks (Fuld, Katzman, Davies, & Terry, 1982; Mohs & Davis, 1982). Mohs and Davis's Alzheimer's disease patients did demonstrate a reduction in liberality of bias with physostigmine treatment (which enhances synaptic acetylcholine availability) evident in both C and B_r as calculated by the present authors. Their experimental design does not allow one to say that this treatment yielded a near-normal pattern of bias because there were no age-matched normal control subjects, but the changes seen were in the desired direction.

However, abnormally liberal bias is evident as well in the Butters' et al. sample of Huntington's disease patients in whom cortical cholinergic deficits are not prominent (Pearce, Sofroniew, Cuello, Powell, Eckenstein, Esiri, & Davison, 1984). Thus, although the dementia deficit in bias is clear, its neurochemistry remains to be elucidated.

References

American Psychiatric Association. (1980). *Diagnostic and statistical manual of mental disorders* (3rd ed.). Washington, DC: Author.

Banks, W. P. (1970). Signal detection theory and human memory. *Psychological Bulletin, 74*, 81–99.

Bartus, R. T., Dean, R. L., Beer, B., & Lippa, A. S. (1982). The cholinergic hypothesis of geriatric memory dysfunction. *Science, 217*, 408–417.

Battig, W. F., & Montague, W. E. (1969). Category norms for verbal items in fifty-six categories: a replication and extension of the Connecticut category norms. *Journal of Experimental Psychology Monograph, 80*(3, Pt. 4).

Bayles, K. A. (1982). Language function in senile dementia. *Brain and Language, 16*, 265–280.

Branconnier, R. J., Cole, J. O., Spera, K. F., & De Vitt, D. R. (1982). Recall and recognition as diagnostic indices of malignant memory loss in senile dementia: A Bayesian analysis. *Experimental Aging Research, 8*, 189–193.

Brinkman, S. D., & Gershon, S. (1983). Measurement of cholinergic drug effects on memory in Alzheimer's disease. *Neurobiology of Aging, 4,* 139–145.

Butters, N., Wolfe, J., Martone, M., Granholm, E., & Cermak, L. S. (1985). Memory disorders associated with Huntington's disease: Verbal recall, verbal recognition and procedural memory. *Neuropsychologia, 23,* 729–743.

Crook, T., Gilbert, J. G., & Ferris, S. (1980). Operationalizing memory impairment for elderly persons: The Guild Memory Test. *Psychological Reports, 47,* 1316–1318.

Egan, J. P. (1958). *Recognition memory and the operating characteristic.* (Tech. Note No. AFCRC-TN-58-51, AO-152650). Bloomington, IN: Indiana University Hearing and Communication Laboratory.

Fuld, P. A., Katzman, R., Davies, P., & Terry, R. D. (1982). Intrusions as a sign of Alzheimer's disease: Chemical and pathological verification. *Annals of Neurology, 11,* 155–159.

Gilund, G., & Shiffrin, R. M. (1984). A retrieval model for both recognition and recall. *Psychological Review, 91,* 1–67.

Glanzer, M., & Adams, J. (1985). The mirror effect in recognition memory. *Memory & Cognition, 13,* 8–20.

Green, D. M., & Moses, F. L. (1966). On the equivalence of two recognition measures of short-term memory. *Psychological Bulletin, 66,* 228–234.

Green, D. M., & Swets, J. A. (1974). *Signal detection theory and psychophysics.* Huntington, NY: Krieger.

Grier, J. B. (1971). Nonparametric indexes for sensitivity and bias: Computing formulas. *Psychological Bulletin, 75,* 424–429.

Grober, E., Buschke, H., Kawas, C., & Fuld, P. (1985). Impaired ranking of semantic attributes in dementia. *Brain and Language, 26,* 276–286.

Hart, S. A., Smith, C. M., & Swash, M. (1985). Recognition memory in Alzheimer's disease. *Neurobiology of Aging, 6,* 287–292.

Hodos, W. (1970). A nonparametric index of response bias for use in detection and recognition experiments. *Psychological Bulletin, 74,* 351–354.

Lezak, M. (1983). *Neuropsychological assessment* (2nd Ed.). New York: Oxford.

Link, S. W., & Heath, R. A. (1975). A sequential theory of psychological discrimination. *Psychometrika, 40,* 77–105.

Lockhart, R. S., & Murdock, B. B., Jr. (1970). Memory and the theory of signal detection. *Psychological Bulletin, 74,* 100–109.

Luce, R. D. (1959). *Individual choice behavior.* New York: Wiley.

Luce, R. D. (1963). Detection and recognition. In R. D. Luce, R. R. Bush, & E. Galanter (Eds.), *Handbook of mathematical psychology* (Vol. 1, pp. 103–189). New York: Wiley.

Martin, A., & Fedio, P. (1983). Word production and comprehension in Alzheimer's disease: The breakdown of semantic knowledge. *Brain and Language, 19,* 124–141.

McKhann, G., Drachman, D., Folstein, M., & Katzman, R. (1984). Clinical diagnosis of Alzheimer's disease: Report of the NINCDS-ADRDA work group. *Neurology, 34,* 939–944.

Mohs, R. C., & Davis, K. L. (1982). A signal detectability of the effect of physostigmine on memory in patients with Alzheimer's disease. *Neurobiology of Aging, 3,* 105–110.

Moskovitch, M. (1982). Multiple dissociations of function in amnesia. In L. S. Cermak (Ed.) *Human memory and amnesia* (pp. 337–370). Hillsdale, NJ: Erlbaum.

Murdock, B. B., Jr. (1974). *Human memory: Theory and data.* Potomac, MD: Erlbaum.

Nebes, R. D., Boller, F., & Holland, A. (1986). Use of semantic context by patients with Alzheimer's disease. *Psychology and Aging, 3,* 261–269.

Noreen, D. L. (1977, August). *Relations among some models of choice.* Paper presented at the Tenth Annual Mathematical Psychology Meetings, University of California, Los Angeles, California.

Ober, B. A., Dronkers, N. F., Koss, E., Delis, D. C., & Friedland, R. P. (1986). Retrieval from semantic memory in Alzheimer-type dementia. *Journal of Clinical and Experimental Neuropsychology, 8,* 75–92.

Paivio, A., Yuille, J. C., & Madigan, S. (1968). Concreteness, imagery, and meaningfulness values for 925 nouns. *Journal of Experimental Psychology Monograph, 76*(1, Pt. 2).

Parkin, A. (1984). Amnesic syndrome: A lesion-specific disorder? *Cortex, 20,* 479–508.

Pearce, R. C. A., Sofroniew, M. V., Cuello, A. C., Powell, T. P. S., Eckenstein, R., Esiri, M. M., & Davison, A. N. (1984). Neurotransmitter dysfunction and atrophy of the caudate nucleus in Alzheimer's disease. *Neurochemical Pathology, 2,* 221–233.

Pollack, I., & Norman, D. A. (1964). A nonparametric analysis of recognition experiments. *Psychonomic Science, 1,* 125–126.

Reisberg, B., Ferris, S. H., & Crook, T. (1982). Signs, symptoms, and course of age-related cognitive decline. In S. Corkin, K. L. Davis, J. H. Growdon, E. Usdin, & R. J. Wurtman. (Eds.), *Alzheimer's disease: A report of progress in research* (Vol. 19, pp. 177–181). New York: Raven Press.

Semple, S. A., Smith, C. A., & Swash, M. (1982). The Alzheimer disease syndrome. In S. Corkin, K. L. Davis, J. H. Growdon, E. Usdin, & R. J. Wurtman (Eds.) *Alzheimer's disease: A report of progress in research* (Vol. 19, pp. 93–108). New York: Raven Press.

Snodgrass, J. G., & Vanderwart, M. (1980). A standardized set of 260 pictures: Norms for name agreement, familiarity and visual complexity. *Journal of Experimental Psychology: Human Learning and Memory, 6,* 174–215.

Thorndike, E. L., & Lorge, I. (1944). *The teacher's word book of 30,000 words.* New York: Columbia University, Teacher's College.

Tulving, E., & Thomson, D. M. (1971). Retrieval processes in recognition memory. Effects of associative context. *Journal of Experimental Psychology, 87,* 116–124.

Upton, G. J. G. (1978). *The analysis of cross-tabulated data.* Chichester, England: Wiley.

Warrington, E. K. (1975). The selective impairment of semantic memory. *Quarterly Journal of Experimental Psychology, 27,* 635–657.

Woodworth, R. S. (1938). *Experimental psychology.* New York: Holt.

(Appendix follows on next page)

Appendix

Summary of Discrimination and Bias Measures

A. SDT with normal distributions

$$d' = z_{FA} - z_H \tag{1}$$

$$\ln(\beta) = \ln[f_o(z_H)/f_n(z_{FA})] \tag{2}$$

$$C = z_{FA} - d'/2 = 0.5(z_{FA} + z_H) \tag{3}$$

B. SDT with logistic distributions

$$d_L = \ln\{[H(1 - FA)]/[(1 - H)FA]\} \tag{4}$$

$$\ln(\beta_L) = \ln\{[H(1 - H)]/[FA(1 - FA)]\} \tag{5}$$

$$C_L = 0.5 [\ln\{[(1 - FA)(1 - H)]/[(H)(FA)]\}] \tag{6}$$

C. Two-high-threshold theory

$$P_r = H - FA \tag{7}$$

$$B_r = FA/[1 - (H - FA)] \tag{8}$$

D. Distribution free (nonparametric) theory

For $H \geq FA$: $A' = 0.5 + [(H - FA)(1 + H - FA)]/[4H(1 - FA)]$ (9A)

For $FA > H$: $A' = 0.5 - [(FA - H)(1 + FA - H)]/[4FA(1 - H)]$ (9B)

For $H \geq FA$: $B'' = [H(1 - H) - FA(1 - FA)]/[H(1 - H) + FA(1 - FA)]$ (10A)

For $FA > H$: $B'' = [FA(1 - FA) - H(1 - H)]/[FA(1 - FA) + H(1 - H)]$ (10B)

For $H \leq (1 - FA)$: $B'_H = 1 - \{[FA(1 - FA)]/[H(1 - H)]\}$ (11A)

For $H > (1 - FA)$: $B'_H = \{[H(1 - H)]/[FA(1 - FA)]\} - 1$ (11B)

Note. $H = (\#\text{hits} + 0.5)/(\#\text{olds} + 1)$; $FA = (\#FAs + 0.5)/(\#\text{olds} + 1)$

Correction to Wolf and Algom (1987)

In the article, "Perceptual and Memorial Constructs in Children's Judgments of Quantity: A Law of Across-Representation Invariance," by Yuval Wolf and Daniel Algom (*Journal of Experimental Psychology: General*, 1987, Vol. 116, No. 4, pp. 381–397), the following sentence was printed incorrectly (p. 381, right-hand column, line 15):

The procedure involves the parallel, though separate, examination of the relations between perceptual stimulus magnitude and memory-based magnitude judgments and between stimulus magnitude and regular judgments.

The sentence should have read as follows:

The procedure involves the parallel, though separate, examination of the relations between stimulus magnitude and memory-based magnitude judgments and between stimulus magnitude and regular perceptual judgments.