Journal of Experimental Psychology:
Learning, Memory, and Cognition
1993, Vol. 19, No. 3, 638–652

0278-7393/93/$3.00

Speed Versus Accuracy Instructions, Study Time, and the Mirror Effect

Kisok Kim and Murray Glanzer

Attention/likelihood theory is a model of recognition memory designed to explain the mirror effect (Glanzer & Adams, 1985, 1990). The theory and the effect were studied using speed versus accuracy instructions and short versus long exposure of stimuli. Speed versus accuracy instructions during test and short versus long exposure of stimuli during study were used to vary the number of features sampled from stimuli. When the number of features sampled was reduced either by speed instructions or by shorter exposure, recognition performance was impaired. The theory predicts that in such cases, all distances between underlying distributions will contract. That means, moreover, that when recognition accuracy is decreased for old stimuli, it is also decreased for new stimuli. These predictions were supported by the data from three experiments.

In two studies, we examined the mirror effect in recognition memory and attention/likelihood theory, a model for recognition memory designed to explain the mirror effect (Glanzer & Adams, 1985, 1990). Two extensions of the theory were developed and tested.

The theory incorporates a signal-detection theory analysis of memory (Egan, 1958). In such an analysis, old and new stimuli in a memory test form distributions as depicted in Figure 1. If there are two classes of stimuli, then two pairs of distributions have to be considered. Figure 2 shows the possible arrangement of the two pairs of distributions when subjects' recognition performance is better with Class A stimuli than Class B stimuli. In such a case, there are several patterns of underlying distributions that could give differences in recognition performance.

Panel 2 of Figure 2 represents the situation in which performance is better with Class A than Class B because subjects recognize new A stimuli better than new B stimuli. Old A stimuli and old B stimuli, however, do not show a difference in recognition. Panel 3 of Figure 2 represents the situation in which the difference in performance is produced because subjects recognize old A stimuli better than old B stimuli. In this case, however, new A stimuli and new B stimuli do not show a difference in recognition.

Panel 1 of Figure 2 represents the situation in which subjects recognize old Class A stimuli better than old Class B stimuli and also recognize new Class A stimuli better than new Class B stimuli. It turns out that this arrangement of distributions is generally the case when one class of stimuli is recognized better than the other (Glanzer & Adams, 1985). Because the order of the distributions of the old stimuli is the reverse of the order of the distributions of the new stimuli, the arrangement is called the *mirror effect*. The definition of the mirror effect is as follows: If there are two classes of stimuli, A and B, and if one class of stimuli, A, is recognized better than another class of stimuli, B, then two things hold: Class A is recognized better than Class B as old when old, and Class A is recognized better than Class B as new when new.

Recognition memory with word frequency as a variable gives one example of the mirror effect. Low-frequency words are recognized better than high-frequency words. Because the mirror effect holds, low-frequency old words are recognized better than high-frequency old words as old and also low-frequency new words are recognized better than high-frequency new words as new. A meta-analysis of 80 recognition memory experiments (Glanzer & Adams, 1985) that included many different variables in addition to word frequency, such as concreteness and pictures versus words, has demonstrated the generality of the mirror effect.

ATTENTION/LIKELIHOOD THEORY

Attention/likelihood theory (Glanzer & Adams, 1990) has been designed to explain the mirror effect. It is a feature-sampling theory, closely related to stimulus-sampling theory (Bower, 1972; Estes, 1950, 1955), that incorporates the mechanisms of signal-detection theory. In particular, this theory adopts completely the following idea from signal-detection theory: Subjects make recognition decisions by considering a test item's likelihood ratio, its probability of being old versus new.

In two studies, we extended attention/likelihood theory by developing a more complete picture of the role of feature sampling. The basic theory consists of the following assumptions.

1. Stimuli are sets of features. The number of such features is N. This is assumed to be constant for all stimuli. Each feature is in one of two states. It is either marked as old (i.e., as experienced on the study list) or unmarked. This applies to both old and new stimuli, as noted later.

This article was based on a doctoral thesis by Kisok Kim. The research was supported by Grant 1 RO1 MH449 from the National Institute of Mental Health and Grant BNS 84–15904 from the National Science Foundation.

We thank John K. Adams, Martin Braine, Lloyd Kaufman, Larry Maloney, and Gay Snodgrass for their help in the development of the research. We also thank Malvina Donskaya for her help in collecting data.

Correspondence concerning this article should be addressed to either Kisok Kim or Murray Glanzer, Department of Psychology, 8th Floor, New York University, 6 Washington Place, New York, New York 10003.

2. Some proportion of features, p(new), is already marked in new stimuli. This parameter represents the noise level, the number of marked features in a new stimulus. It is assumed to be constant for all stimuli because there is no reason to assume, at this point, that one stimulus enters with greater noise marking than another.

3. Different classes of stimuli evoke different amounts of attention by the subject. This is represented by differences in the number of features, $n(i)$, sampled during examination. The i refers to a class of stimuli.[1] The sampling of features for a given stimulus is assumed to be random.

4. During study, all of the $n(i)$ features sampled are marked if they are not already marked. The proportion of features sampled is $\alpha(i) = n(i)/N$. Because some proportion of features, p(new), is already marked before study (see Assumption 2), the state of stimuli in class i, after study, is given by the following equation:

$$p(i, \text{old}) = p(\text{new}) + \alpha(i) \cdot [1 - p(\text{new})]. \qquad (1)$$

This states that the subject samples a proportion, $\alpha(i)$, of the features. Those already marked, p(new), remain marked; those not yet marked, $1 - p$(new), are now marked. Classes of stimuli that evoke examination of a larger proportion of features result in the marking of a larger proportion of features; the learning constant $\alpha(i)$ is larger.

5. During a recognition test, a subject also samples a set of features, $n(i)$, from a stimulus. Because the sampling is random, the features sampled during test are, for an old stimulus, not necessarily the same as those sampled during study. It follows from the preceding statements that the number of marked features is binomially distributed with parameters $n(i)$ and $p(i$, old) for old stimuli, and parameters $n(i)$ and p(new) for new stimuli.

6. During the test, the subject checks each sampled feature to see whether it is marked. The subject counts the number of features that are marked. The subject then estimates, on the basis of experience with such an item during the study period, what marking an old item is likely to have and what marking a new item is likely to have. These estimates are used to compute a likelihood ratio for that item, as presented in Equation 2. Decisions are based on the likelihood ratio.

To recapitulate, during study a number of features, $n(i)$, are sampled and those still unmarked are marked. After study, all of the $n(i)$ sampled features are marked and those marked features are added to the set of unsampled marked features.

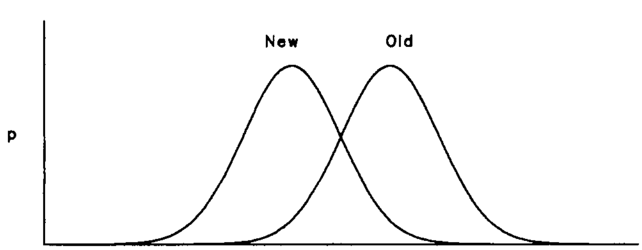

Figure 1. Representation of new and old stimuli in a recognition memory test according to signal-detection theory.

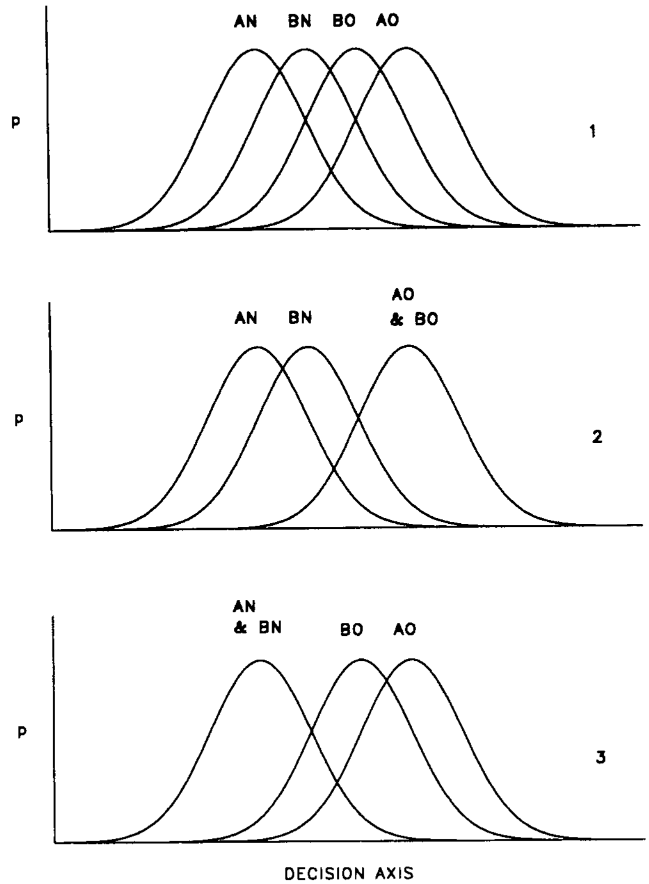

Figure 2. Three possible orders of underlying distributions when accuracy for Class A was greater than accuracy for Class B. (A = Class A, B = Class B, O = old, and N = new. Panel 1 shows the mirror effect.)

During test, $n(i)$ features are again sampled both for new and old stimuli. In this sample of $n(i)$ features, the subject finds a number of already marked features, x. The difference between $n(i)$ and x should be noted. The $n(i)$ is the sample size, with x features marked and the remainder, $n(i) - x$, unmarked. The subject then uses the processes prescribed by signal-detection theory in making responses. Likelihood ratios are computed (see Equation 2) and decisions are made on the basis of those likelihood ratios. The likelihood ratios are computed for every stimulus item and are the basis of the decision made about every item. The generation of likelihood ratios for the stimulus presented and their use for the recognition decision is what produces the mirror effect. The use of likelihood ratios rather than strength measures makes the theory different from other theories of recognition memory.

[1] In this article we use i to refer to stimulus class. However, Glanzer and Adams (1990) pointed out that other variables (e.g., encoding task) can have the same function as stimulus class. This assumption could therefore be worded in a more general form.

Illustrative Computation

To clarify the theory, we offer a simple computation. Assume that the total number of features in any stimulus is $N = 1{,}000$, the number of features sampled in a stimulus in Class A is $n(A) = 81$, the number of features sampled in a stimulus in Class B is $n(B) = 63$, and $p(\text{new}) = .15$. From these parameters, $p((A, \text{old})$ and $p(B, \text{old})$ are computed using Equation 1. Therefore, $p(A, \text{old}) = .22$ and $p(B, \text{old}) = .20$.

During the recognition test, as during study, the subject randomly samples 81 features from Class A stimuli and 63 features from Class B stimuli. The number of marked features among the sampled features is, then, binomially distributed with parameters, $n(A) = 81$ and $p(A, \text{old}) = .22$ for AO; $n(A) = 81$ and $p(\text{new}) = .15$ for AN; $n(B) = 63$ and $p(B, \text{old}) = .20$ for BO; and $n(B) = 63$ and $p(\text{new}) = .15$ for BN. Figure 3 shows the two pairs of binomial distributions for new and old stimuli in Classes A and B.

The log likelihood ratio distributions that are obtained from the binomial distributions of Figure 3 are shown in Figure 4. To construct the distributions in Figure 4, two steps are carried out. The log likelihood ratio is computed for every x in the Class A distributions and every x in the Class B distributions. These two sets of log likelihood ratios are then plotted on a single likelihood axis, distributed according to the probabilities of the distributions that originally produced them: AN, AO, BN, BO. The log likelihood ratios,[2] $\lambda(x)$, for the number of marked features, x, for items in each stimulus class are computed as follows. In the equation, x is the number of marked features, i is the stimulus class (e.g., high- or low-frequency words), and q is the complement of p.

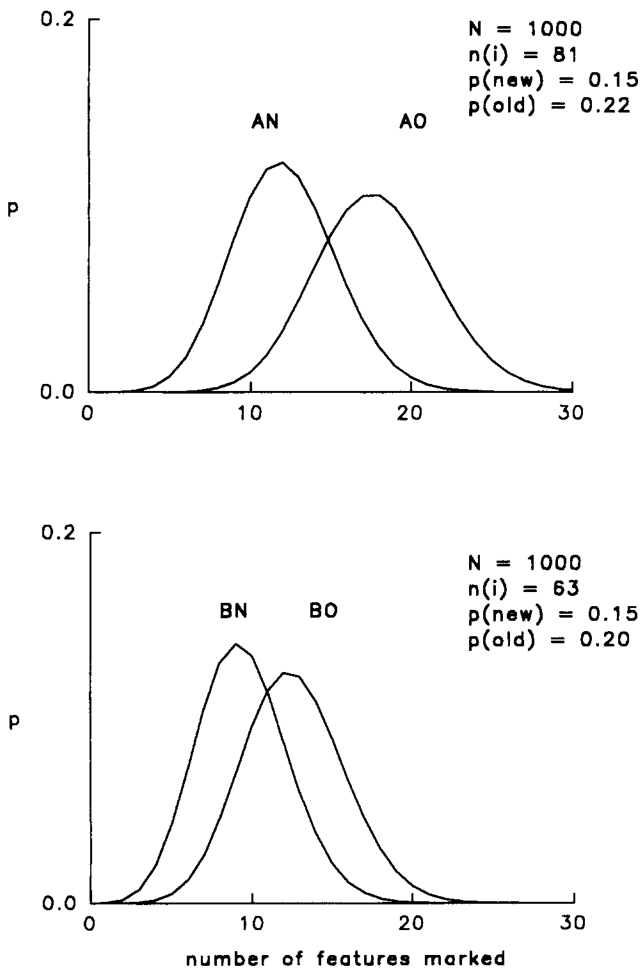

Figure 3. The distributions of number of features marked for Class A (top) and Class B (bottom) stimuli. (A = Class A, B = Class B, N = new, and O = old. They are binomially distributed with the parameters shown.)

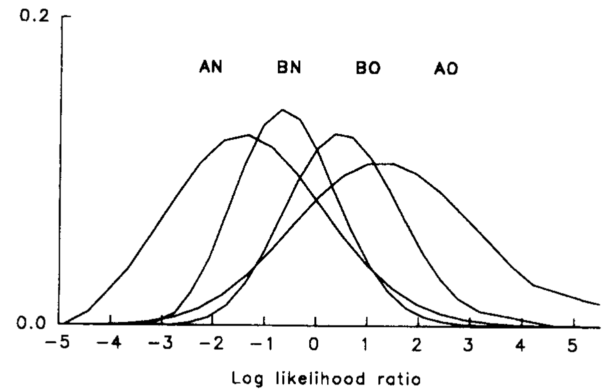

Figure 4. The log likelihood ratio distributions constructed from the binomial distributions in Figure 3. (A = Class A, Class B, N = new, and O = old.)

$$\lambda(x) = \ln\left[\frac{\binom{n(i)}{x}\cdot p(i,\text{old})^x\cdot q(i,\text{old})^{n(i)-x}}{\binom{n(i)}{x}\cdot p(\text{new})^x\cdot q(\text{new})^{n(i)-x}}\right]$$

$$= \ln\left[\frac{p(i,\text{old})^x\cdot q(i,\text{old})^{n(i)-x}}{p(\text{new})^x\cdot q(\text{new})^{n(i)-x}}\right] \tag{2}$$

$$= n(i)\cdot\ln\left[\frac{q(i,\text{old})}{q(\text{new})}\right] + x\cdot\ln\left[\frac{p(i,\text{old})\cdot q(\text{new})}{p(\text{new})\cdot q(i,\text{old})}\right].$$

These values of $\lambda(x)$ are binomially distributed according to the following equation:

$$p[\lambda(x)] = \binom{n(i)}{x}\cdot p(i,j)^x\cdot q(i,j)^{n(i)-x}. \tag{3}$$

where i is again stimulus class, j is state of the item (old or new), and $\lambda(x)$ is log likelihood ratio as given by Equation 2. It is important to note that the binomial distribution enters twice. First, it is used in computing the likelihood ratios of Equation 2. Second, it is again used to distribute those likelihood ratios (see Equation 3) with a different

[2] The term $\lambda(x)$ in Equation 2 should more properly be written as $\lambda(x \mid i)$ and the term $p[\lambda(x)]$ in Equation 3 written as $p[\lambda(x \mid i) \mid j]$. We have used the abbreviated forms $\lambda(x)$ and $p[\lambda(x)]$ to simplify the presentation.

distribution for each combination of i and j. Note also that even though p(new) is assumed to be the same for two classes of stimuli (such as Classes A and B), when distributions of likelihood ratios are plotted, the AN and BN distributions of likelihood ratios separate to produce the mirror order as in Figure 4.

The generation of a likelihood ratio for every stimulus means that when subjects make a recognition decision, they consider the probability of each test item's being old and that of its being new at the same time. Therefore, the recognition decision is based on the comparison of the two probabilities. This combining of information about old and new items in every decision is what produces the mirror effect.

Attention/Likelihood Theory in Two-Alternative Forced Choice

We describe in detail how attention/likelihood theory works in the context of the two-alternative forced-choice test because that test was used in the experiments that follow. We continue with the example we used earlier to illustrate the theory, with two classes of stimuli, A (better recognized) and B. To make the example more concrete, we now use low-frequency words as an example of Class A stimuli and high-frequency words as Class B stimuli.

We used an extended form of the two-alternative forced-choice test. Usually, in a two-alternative forced-choice recognition test, an old stimulus is paired with a new stimulus. Ordinarily, when two classes of stimuli (e.g., high- and low-frequency words) are used, there are four standard forced-choice conditions, as follows: HO versus HN, LO versus HN, HO versus LN, and LO versus LN. H stands for high frequency, L for low frequency, O for old, and N for new. We, however, used two additional forced choices, as in Glanzer and Bowles (1976), called *null choice conditions*: HN versus LN and LO versus HO. They are called the null choice conditions because in the case of HN versus LN, both members of the pair are new and, in the case of LO versus HO, both members of the pair are old. We discuss the relevance of the null choice conditions to the mirror effect shortly.

When the six types of choices are used in a recognition test, the following pattern of proportions of choice is expected with the mirror effect. In standard pairs,

$$P(\text{HO, HN}) < P(\text{LO, HN}), P(\text{HO, LN}) < P(\text{LO, LN});$$

in null pairs,

$$P(\text{HN, LN}) > .50, \qquad P(\text{LO, HO}) > .50.$$

Each expression represents the proportion of choice of the first term in the parentheses over the second term. For example, P(HO, HN) stands for the proportion of choice of HO over HN. The pattern of inequalities corresponds to the underlying distributions in the mirror pattern shown in Figure 4. The distances between distributions in Figure 4 determine the order of proportions of choice. For example, the distance between LO and LN is the largest in Figure 4. Therefore, P(LO, LN) is the largest in the set of inequalities for the standard pairs. The distance between HO and HN is the smallest of the standard pairs; therefore, P(HO, HN) is the smallest. The null choice conditions give direct measures of the distance between the two new distributions (HN and LN) and between the two old distributions (LO and HO). The expected pattern of results for both standard and null pairs has been shown in previous studies (Bowles & Glanzer, 1983; Glanzer, Adams, & Iverson, 1991; Glanzer & Bowles, 1976).

Attention/likelihood theory predicts the proportions of choices themselves as well as the order of proportions, provided that appropriate parameters are fixed. The following shows how the proportions are computed with the parameters we used earlier: $N = 1{,}000$, $n(\text{high}) = 63$, $n(\text{low}) = 81$, and $p(\text{new}) = .15$. (Steps 1 and 2 have been described earlier and are not repeated in detail here.)

1. Two pairs of binomial distributions are constructed for high- and low-frequency word stimuli. Figure 3 shows the binomial distributions.

2. Likelihood ratio distributions are constructed from the binomial distributions using Equations 2 and 3. Figure 4 shows the likelihood ratio distributions.

3. Six bivariate distributions of likelihood ratios (HO × HN, LO × HN, HO × LN, LO × LN, HN × LN, and LO × HO) are constructed with probability function $f[\lambda(x), \lambda(x')]$, as follows:

$$\begin{aligned} f[\lambda(x), \lambda(x')] &= p[\lambda(x)] \cdot p[\lambda(x')] \\ &= \binom{n(i)}{x} \cdot p(i,j)^{x} \cdot q(i,j)^{n(i)-x} \qquad (4) \\ &\quad \cdot \binom{n(i')}{x'} \cdot p(i',j')^{x'} \cdot q(i',j')^{n(i')-x'}. \end{aligned}$$

The equation is the product of the two binomially distributed probabilities. $\lambda(x)$ and $\lambda(x')$ represent the log likelihood ratio of each term (e.g., $\lambda[x]$ for HO and $\lambda[x']$ for LN in HO × LN distribution). As in Equation 2, i and i' refer to classes of stimuli (e.g., high- or low-frequency words), j and j' refer to old or new states, and x and x' refer to the number of features marked in states i, j and i', j'. For example, if the test pair is LO versus HN, i = low-frequency word, j = old, i' = high-frequency word, and j' = new. The meaning of Equation 4 can be more easily understood in Step 4.

4. Given a pair of stimuli (e.g., LO vs. HN), the expected proportions of choice of the first term (LO) is the sum of the probabilities (from Equation 4) of all pairs of likelihood ratios for which the likelihood ratio for one of the conditions (LO) is greater than for the other (HN). Ties in the likelihood ratio are handled by assigning half of the probabilities to each condition. Each of the six bivariate distributions is used for each of the six types of choice pairs.

Given this procedure and the parameters $N = 1{,}000$, $n(\text{high}) = 63$, $n(\text{low}) = 81$, $p(\text{new}) = .15$ that were used in the example presented earlier, the expected proportions are the following. For the standard choices, P(HO, HN) = .78, P(LO, HN) = .82, P(HO, LN) = .83, and P(LO, LN) = .86. For the null choices, P(HN, LN) = .64 and P(LO, HO) = .62. Note that these proportions show the mirror effect because for standard choices, P(HO, HN) < P(LO, HN),

$P(\mathrm{HO}, \mathrm{LN}) < P(\mathrm{LO}, \mathrm{LN})$; for the null conditions, $P(\mathrm{HN}, \mathrm{LN}) > .5$, $P(\mathrm{LO}, \mathrm{HO}) > .5$.

Effect of Varying $n(i)$

Up to this point, $n(i)$, the attention parameter, has been considered constant for a given stimulus class, i, over both study and test. In this section, and the sections that follow, we introduce variations in procedure designed to vary $n(i)$. These variations will, on the basis of an extension of the theory, produce predicted changes in the pattern of choices.

In particular, we introduce operations designed to change $n(i)$ during test in one case and during study in the other. In Experiment 1, we varied $n(i)$ during test by speed versus accuracy instructions. In Experiments 2a and 2b, we varied $n(i)$ during study by varying study time.

Attention/likelihood theory predicts that recognition performance should decline with decreased $n(i)$ and that the decline should show in responses to new stimuli as well as in those to old stimuli. The theory, more generally, predicts a systematic pattern of changes in the mirror effect; this pattern of changes is represented in the distributions of Figure 5. New distributions as well as old distributions move toward the center of the array (i.e., the change in the positions of the new distributions mirrors that of the old distributions). These movements will be evidenced by the systematic changes in the proportions of choice in the forced-choice tests.

We report two studies that tested the theory's predictions.

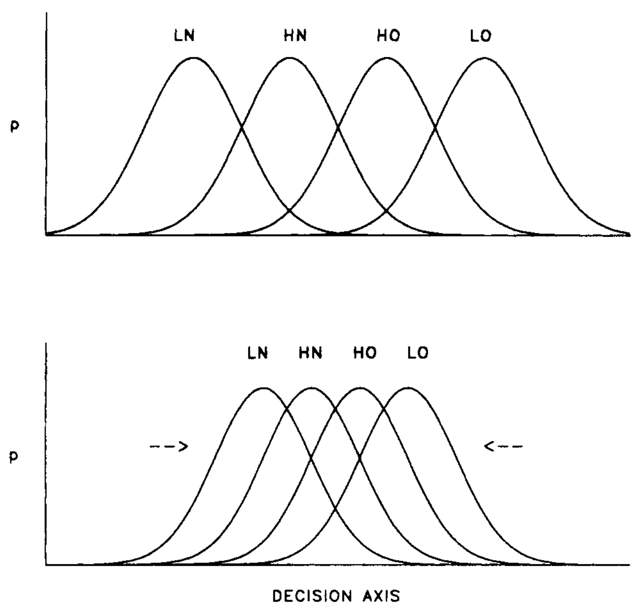

Figure 5. The general pattern of changes in the mirror effect with reduction of $n(i)$, the attention parameter. (Word frequency was the variable used to produce the mirror effect. The bottom picture shows the concentered distributions produced by speed instructions or by short study time. LO = low-frequency old, HO = high-frequency old, HN = high-frequency new, and LN = low-frequency new. This is a generalized, schematic representation. Distributions computed on the basis of the theory can be found in Figures 6 and 7.)

In each experiment, we used the forced-choice test with word frequency as a secondary independent variable. In Experiment 1, speed versus accuracy instructions were given during the recognition test to vary $n(i)$ during the test. In Experiments 2a and 2b, short versus long exposure of stimuli was given during list presentation to vary $n(i)$ during study.

Experiment 1: Varying n(i) During Test—Speed Versus Accuracy Instructions

In previous applications of the theory, it has been assumed that the attention parameter, referred to as $n(i)$ for a given class of stimuli i, is the same during study and test (Glanzer & Adams, 1990; Glanzer et al., 1991). We now extend the theory to cover the effect of speed versus accuracy instructions by asserting that the instruction variable will change the attention parameter—the number of features sampled—during test. Under speed instructions that parameter will be reduced during test. In order to refer to the several possible states of $n(i)$ under discussion, we increase the number of arguments in the term. It is now written as $n(i, k, l)$, where i refers as before to high- or low-frequency words; k refers to point in time, study, or test; and l refers to instruction condition, speed, or accuracy.

Because the speed versus accuracy instruction is given after study is completed, the attention parameters for speed and accuracy conditions during study, $n(i, k, l)$, are equivalent and are expressed as follows: $n(i, \text{study}, \text{speed}) = n(i, \text{study}, \text{accuracy})$. Because these two are equal, we refer to them jointly as $n(i, \text{study}, .)$ Attention, as indexed by $n(i, k, l)$, is also assumed to remain constant from study to test in the accuracy condition. Therefore, $n(i, \text{test}, \text{accuracy}) = n(i, \text{study}, .)$. Attention is, however, reduced by speed instructions during test so that $n(i, \text{test}, \text{speed}) = \delta \cdot n(i, \text{study}, .)$, where δ is the coefficient of reduction of stimulus sampling during test and $0 < \delta < 1$. In other words, the attention parameter is reduced by the factor δ when speed instructions are given during test.

As stated earlier, attention/likelihood theory predicts that recognition will be worse with decreased $n(i, k, l)$ and that the decline in recognition performance should show in poorer recognition of new stimuli as well as in poorer recognition of old stimuli. In terms of the underlying distributions, new distributions as well as old distributions will move toward the center of the distributions, as in Figure 5. This will be observed in the decline of all six proportions of choice with speed instructions. The decline of the proportions is the evidence for the decrease in distances between underlying distributions.

Method

Subjects were given a lexical decision task that was followed by a recognition memory test for the words shown in the lexical decision task. The lexical decision task was used as an encoding task. Its purpose was to have subjects process each item. During the recognition test, speed instructions were given to one group of subjects and accuracy instructions to another.

Subjects. The subjects were 48 undergraduate students from an introductory psychology course. They participated in the experiment to fulfill a course requirement. All subjects were native speakers of English. The description provided here of the subjects—source, basis of participation, and language background—is the same for the subjects in Experiments 2a and 2b.

Design. There were two independent variables. The main variable was speed versus accuracy instructions. The other variable was word frequency. We used the word-frequency variable to have two sets of old and new distributions on which the effect of the main variable could be examined. The subjects were randomly assigned to one of two conditions: 24 subjects to the speed condition and remaining 24 to the accuracy condition. The dependent variable was the subjects' recognition memory performance in a two-alternative forced-choice recognition test. Instructions was a between-subjects variable and word frequency was a within-subjects variable.

Procedure. An Apple II computer was used to present stimuli and record responses in both the lexical decision task and the recognition test. The subjects were not told that they were going to have a recognition test until they had completed the lexical decision task.

In the lexical decision task, an equal number of English words and nonwords were presented one at a time in random order. Half of the words were high-frequency words and the other half were low-frequency words. The first and last 10 items on the list were filler items, none of which appeared in the later recognition test. They were included to eliminate primacy and recency effects on the main list items.

The subjects were seated before a monitor and keyboard. During the lexical decision task, one key on the keyboard was labeled *yes* and another was labeled *no*. The two response keys were arranged horizontally next to each other. Subjects were instructed to press the *yes* key when they thought the presented item was a word and the *no* key otherwise. The task was self-paced, with the words and nonwords staying on the monitor until the subjects made a response. Feedback—the words *RIGHT* or *WRONG* displayed on the monitor—was given after each lexical decision response.

After the lexical decision task was completed, the subjects were given instructions for the recognition test that followed. The recognition test included the words they had seen in the lexical decision task and new words. The subjects given speed instructions were told to respond as quickly as possible. The subjects who received accuracy instructions were told to respond as accurately as possible. Feedback was given in each test condition to maintain the speed or the accuracy set. In the speed condition, subjects were given feedback on their response time, in the form of a number ranging from 1 to 9. When a response was made within 750 ms after a test pair was presented, a 1 was displayed. The feedback number was increased by 1 with each 250-ms increment in response time. For example, a response time between 750 ms and 1,000 ms was given a 2. Subjects were told that a 1 or 2 meant a fast response, 3 (1,000–1,250 ms) was acceptable, and number larger than 3 meant a slow response. In the accuracy condition, if a subject's response was correct, the word *RIGHT* was displayed on the monitor; otherwise, *WRONG* was displayed. For null choice test pairs, HN–LN and LO–HO, *RIGHT* feedback was always given because there were no correct or incorrect responses. In both instruction conditions, the feedback instruction was displayed on the monitor immediately after the subject responded.

The recognition task consisted of a forced choice between pairs of words. There were six kinds of test pairs: HN–LN, LO–HO, HO–HN, LO–HN, HO–LN, and LO–LN. Each pair was presented on the monitor, arranged vertically, one word above the other. There were two response keys on the keyboard, one labeled with an arrow pointing up, the other, below it, with an arrow pointing down. Subjects pressed the key with the upward arrow to choose the upper word as the one seen earlier and the key with the downward arrow to choose the lower word. The stimuli remained on the screen until the subject responded. After each response, the subject saw the appropriate feedback information.

Material. Sets of 180 high-frequency words (40 or more times per million) and 180 low-frequency words (8 or fewer times per million) were selected from Kučera and Francis (1967). Word lengths ranged from 3 to 10. The two groups of words were matched on word length and concreteness value according to Paivio, Yuille, and Madigan's (1968) norms.

Half of the high-frequency words (90), and half of the low-frequency words (90) were presented in the lexical decision task. A set of 180 nonwords was also constructed for the lexical decision task. The study lists were individually randomized for each subject.

In order to make the six kinds of test pairs (HN–LN, LO–HO, HO–HN, LO–HN, HO–LN, and LO–LN), the 180 high-frequency words were divided into six groups, with 30 words in each group, and each group was matched in terms of word frequency, concreteness, and word length. The 180 low-frequency words were divided in the same way into six groups. The six groups of high-frequency words and the six groups of low-frequency words were combined to form the six groups of word pairs, with 30 pairs in each group. Words in each pair were matched in word length. Within each pair condition, the position of words on the monitor (top or bottom) was counterbalanced. Moreover, words used in one condition for one subject (e.g., HN) were used in the other condition (e.g., HO) for another subject. Stimulus words were completely rotated through the experimental conditions across the subjects.

Results and Discussion

Lexical Decision. The results on performance on the lexical decision task are important only as background information. Because the lexical decision task occurred before the speed versus accuracy test instructions, we expected performance on the lexical decision task to be the same across the two experimental conditions, and it was. The mean proportions of correct responses for subjects who were later given speed instructions for the recognition test were as follows: high-frequency words = .99, low-frequency words = .91, and nonwords = .93. The corresponding proportions of correct responses for subjects later given the accuracy instructions were as follows: high-frequency words = .99, low-frequency words = .90, and nonwords = .93.

There was no statistical difference in the proportion of correct lexical decisions between the speed and accuracy conditions, $F(1, 46) = 0.011$, $p = .9167$, $MS_e = 0.0383$. However, the differences between item classes (high-frequency words, low-frequency words, and nonwords) were, as expected, statistically significant, $F(2, 92) = 70.73$, $p < .0001$, $MS_e = 0.041$. Analyses of proportions here and in the rest of this article were carried out on the arcsine transformation of the original proportions.

The response times during lexical decision for the two groups again, as expected, did not differ. The mean response times were as follows: for the speed condition group, the means (in milliseconds) were 697 (SD = 219), 856 (SD = 252), and 980 (SD = 268) for high-frequency words, low-frequency words, and nonwords, respectively. For the accuracy condition group, the means (in milliseconds) were 711 (SD = 172), 890 (SD = 244), and 983 (SD = 295) for high-frequency words, low-frequency words, and nonwords, respectively. There was, as expected, no statistically signif-

Table 1
Experiment 1: Mean Proportions of Choice for Groups Receiving Speed and Accuracy Instructions During Test

	Accuracy condition					
	Null pairs		Standard pairs			
M	*P*(HN, LN)	*P*(LO, HO)	*P*(HO, HN)	*P*(LO, HN)	*P*(HO, LN)	*P*(LO, LN)
Observed	.65	.62	.77	.81	.83	.88
Predicted	.64	.62	.77	.82	.83	.86
	Speed condition					
Observed	.59	.59	.70	.75	.71	.78
Predicted	.60	.60	.70	.75	.75	.78

Note. There were 24 subjects in each group. H = high-frequency words, L = low-frequency words, N = new, and O = old. Parameters: $p(\text{new}) = .1$, $N = 1{,}000$, $n(\text{high, study},.) = 53$, $n(\text{low, study},.) = 69$, and $\delta = 0.527$.

icant difference between speed versus accuracy instruction conditions, $F(1, 46) = 0.029$, $p = .8663$, $MS_e = 0.0583$. The effect of item class was again statistically significant, $F(2, 92) = 84.04$, $p < .0001$, $MS_e = 0.02$. Analyses of response times were carried out on the logarithms of the response times.

The results of the lexical decision task for response times followed the usual pattern of results found in studies of word frequency (Glanzer & Adams, 1990; Glanzer & Ehrenreich, 1979; Scarborough, Cortese, & Scarborough, 1977). Response times were fastest for high-frequency words, slower in low-frequency words, and slowest for nonwords. On accuracy, the subjects were more accurate on high-frequency words than low-frequency words. The greater accuracy on nonwords than low-frequency words does not agree with some of the findings in the literature (Glanzer & Ehrenreich, 1979) but does agree with others (Forster & Chambers, 1973). Lexical decision performance on nonwords depends on how they were constructed and therefore varies in different studies. The important information, however, is that before being given speed versus accuracy instructions, the two experimental groups show matched performance.

Recognition Test. We excluded the words that were responded to incorrectly as nonwords in the lexical decision task when recognition test data were analyzed. Six percent of the accuracy condition items and 6% of the speed condition items were excluded on this basis. Inclusion of those responses, however, did not affect the general pattern of results or their statistical analysis.[3]

The mean proportions of choices in each of the six forced-choice recognition conditions for the accuracy and the speed conditions are presented in the "Observed" rows in Table 1. These proportions are the averages for 24 subjects each in the speed and the accuracy conditions.

It is important to note that the mirror order was present in both the accuracy and speed conditions. As discussed in the introduction, the expected mirror pattern in the forced-choice conditions was, in the standard conditions, $P(\text{HO, HN}) < P(\text{LO, HN})$, $P(\text{HO, LN}) < P(\text{LO, LN})$, and in the null conditions, $P(\text{HN, LN}) > .5$ and $P(\text{LO, HO}) > .5$. Examination of Table 1 shows that all expected relations are present.

Of major concern was the effects of speed versus accuracy instructions. According to the theory, all proportions of choices in the accuracy condition should decrease in the speed condition. Examination of Table 1 shows that those decreases occurred. The decreases corresponded to the predicted concentering of underlying distributions as shown in Figure 5.

In order to analyze the data, we converted the proportions of choices to arcsines and subjected them to a two-way analysis of variance. The main effect of interest was that for speed versus accuracy instruction. The effect was statistically significant, $F(1, 46) = 14.21$, $p < .0005$, $MS_e = 0.16$. There were, of course, statistically significant differences between the six choice conditions, $F(5, 230) = 42.75$, $p < .0001$, $MS_e = 0.053$, that reflected the mirror inequalities. There was no significant interaction between choice type and the instruction variable, $F(5, 230) = 2.115$, $p > .05$, $MS_e = 0.053$.

To determine whether the mirror inequalities held, we carried out two additional types of statistical tests. To check whether the inequalities for the four standard choice conditions—$P(\text{HO, HN}) < P(\text{LO, HN})$, $P(\text{HO, LN}) < P(\text{LO, LN})$—held, we subjected these proportions of choices to linear trend analysis. The results of the analysis were statistically significant, $F(1, 115) = 29.008$, $p < .001$, $MS_e = 0.053$. The proportion of variance accounted for was .99. To check whether $P(\text{LO, HO})$ and $P(\text{HN, LN})$ were greater than .50 as required, we tested the difference of the proportions of choice in the two null choice conditions from .50 with t tests. The results were also statistically significant. In the HN–LN condition, $t(47) = 7.07$, $p < .005$, and in the LO–HO condition, $t(47) = 6.45$, $p < .005$.

We also analyzed the response times during test to determine whether the speed versus accuracy instructions worked as expected, with response times greater in the accuracy condition. Table 2 shows that this was indeed the case. The mean response time in the accuracy condition was almost twice the mean response time in the speed condition. The effect of speed versus accuracy instruction was statistically signifi-

[3] The mean proportions of choice, including items classified incorrectly during lexical decision tasks, were as follows: In the accuracy instruction condition, .65 (HN–LN), .60 (LO–HO), .76 (HO–HN), .78 (LO–HN), .83 (HO–LN), and .87 (LO–LN); in the speed instruction condition, .59 (HN–LN), .59 (LO–HO), .70 (HO–HN), .75 (LO–HN), .71 (HO–LN), and .76 (LO–LN). Compare these with the entries in Table 1. (H = high-frequency words, L = low-frequency words, N = new, and O = old.)

Table 2
Experiment 1: Mean Response Times for Groups Receiving Speed and Accuracy Instructions During Test

	Null pairs		Standard pairs			
Instruction	P(HN, LN)	P(LO, HO)	P(HO, HN)	P(LO, HN)	P(HO, LN)	P(LO, LN)
Accuracy						
M	1,925	1,562	1,631	1,542	1,570	1,508
SD	656	441	523	475	381	358
Speed						
M	1,075	957	985	967	986	981
SD	242	242	182	188	191	179

Note. There were 24 subjects in each group. H = high-frequency words, L = low-frequency words, N = new, and O = old.

cant, $F(1, 46) = 47.385, p < .0001, MS_e = 0.34$. The effect of the six choice types was also statistically significant, $F(5, 230) = 22.92, p < .0001, MS_e = 0.0074$. There was a significant interaction between instruction and choice type, $F(5, 230) = 3.742, p < .003, MS_e = 0.0074$, reflecting a relatively large drop in the response time for HN–LN in the speed condition. This large drop may be related to the fact that the HN–LN condition in the accuracy condition showed the largest response time of all conditions. A similar pattern was observed for this condition in the subsequent experiments.

The response times in the standard choices showed a pattern related to the mirror order of the proportions of choice. They were, as might be expected, related to ease of choices. The easiest choice, LO–LN, the most likely to be correct, tended to be the fastest. The most difficult, HO–HN, tended to be the slowest. The pattern was seen in the averages across the two experimental conditions of this experiment. It was also observed in Experiment 2a (see Table 4) and with a minor deviation in Experiment 2b (see Table 6).

Our next concern was to fit the theory to the data. Table 1 shows the obtained proportions of choices and predicted proportions computed from the theory. The predicted proportions of choice were based on the following parameters: $N = 1{,}000$, $p(\text{new}) = 0.1$, $n(\text{high, study},.) = 53$, $n(\text{low, study},.) = 69$, and $\delta = 0.527$. The fifth parameter, δ, $0 < \delta < 1$, obtains $n(i, \text{test, speed})$ from $n(i, \text{study},.)$ with $n(i, \text{test, speed}) = \delta{\cdot}n((i, \text{study},.)$. (The reduced $n[i, k, l]$ were rounded off to integers.)

Initial attempts to fit several forced-choice experiments showed *N*s close to 1,000, ranging from 993 to 1,024. We therefore fixed *N* at 1,000 for all fits. The remaining parameters were obtained by using PRAXIS, a function minimization algorithm (Brent, 1973) in a fitting program written by Gegenfurtner (1992). The fitting minimized

$$\sum_{m=1}^{12} \frac{[P(m) - \hat{P}(m)]^2}{P(m) \cdot [1 - P(m)]},$$

where m is the choice condition (e.g., comparison of HN and LN in the speed condition), $P(m)$ is the theoretical proportion, $\hat{P}(m)$ is the observed proportion, and $P(m) \cdot [1 - P(m)]$ is the theoretical variance of the predicted proportion in condition m. There were 12 observed and 12 predicted values, 6 each for the speed condition and another 6 each for the accuracy condition. The proportion of variance accounted for by the theory could be estimated by regressing the observed on the predicted values. The R^2 proportion was .98.

Figure 6 shows the theoretical distributions constructed with the estimated parameters. The top figure represents the log likelihood ratio distributions for accuracy condition, and the bottom figure represents those for the speed condition. The figures show that, as predicted by the theory (see Figure

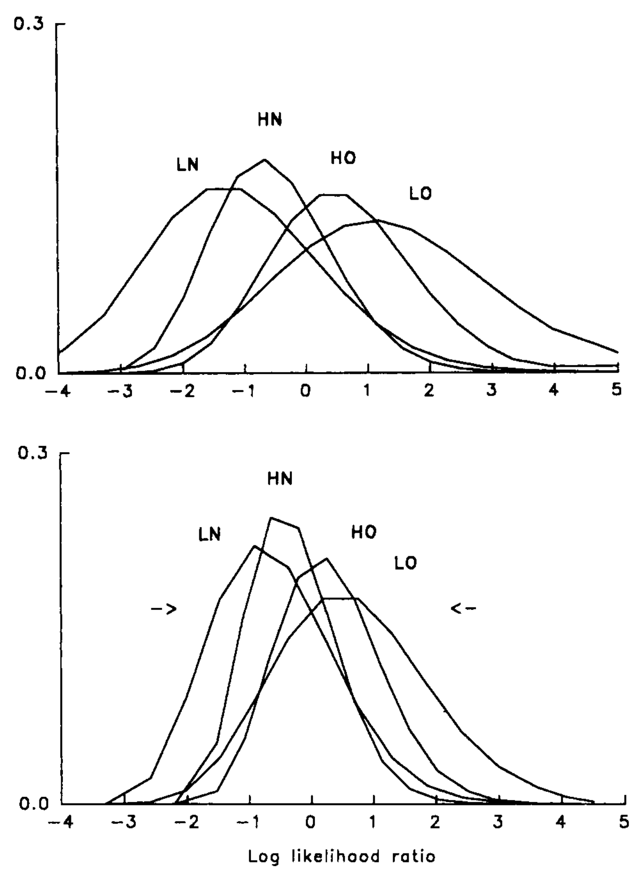

Figure 6. The log likelihood ratio distributions constructed with the parameters obtained by fitting the theory to the data of Experiment 1. (The top panel represents the underlying distributions with accuracy instructions. The bottom panel represents the distributions with speed instructions. LO = low-frequency old, HO = high-frequency old, HN = high-frequency new, LN = low-frequency new.)

5), all distributions in the speed condition moved toward a central point.

Experiment 1 varied the attention parameter, $n(i, k, l)$, during test and produced the predicted pattern of results. We then attempted to carry out a parallel variation of $n(i, k, l)$ during study. The theory predicted, for such a variation, results that parallel those of Experiment 1. Both the good (accuracy) and poor (speed) study conditions should produce a mirror pattern and the poor study condition should produce a concentering of the underlying distributions. Again, the evidence for this concentering should, again, be that all proportions of choice decrease in the poor study condition. We conducted four other studies to demonstrate this systematic change. Two were successful in producing the predicted pattern of results but presented difficulties in interpretation. We summarize them briefly. Two were completely successful and are reported fully.

The first of the two problematic studies was identical to Experiment 1, except that the speed versus accuracy instructions were given during the study phase. The results were as predicted: There was a mirror pattern in the accuracy-during-study condition, a mirror pattern in the speed-during-study condition, and a systematic reduction of all six proportions of choice in the speed condition. Unfortunately, our measures of response times during test indicated a carryover of the speed versus accuracy instructions to the test. The subjects in the speed-during-study condition took less time on the test. Moreover, there was a statistically significant interaction between the carryover with choice conditions; some conditions were affected more than others. This carryover and interaction complicated the interpretation of the results. The effect of the experimental variables could have been caused by changes during study, changes during test, or both.

We restructured the experiment to produce an effect that could be unambiguously assigned to changes in $n(i, k, l)$ during the study phase. In that second study, the speed versus accuracy instruction was again given during study. During test, however, response times were controlled by limiting the exposure time of each test pair and requiring subjects to respond within 0.5 s after each test pair was removed from the monitor. This arrangement was successful in removing the carryover of speed versus accuracy instructions from the study phase. It was also successful in that it showed the predicted pattern of results: a mirror pattern in both the accuracy and speed conditions and a reduction of all proportions of choice in the speed condition. The overall difference between the speed and accuracy conditions was, however, not statistically significant. We therefore decided to use a different operation during the study phase to reduce $n(i, k, l)$ and produce systematic changes in the proportions of choice. We applied this operation in the following two studies.

Experiments 2a and 2b: Varying n(i) During Study—Long- Versus Short-Study Time

In Experiment 1, we demonstrated the effect of varying the attention parameter, $n(i, k, l)$, during test as predicted by the theory. In the next two experiments, we varied $n(i, k, l)$ during the study. We did this by varying study time. The theory predicts that reducing study time will produce the same effect on the underlying distributions: concentering. This would be evidenced in a decline of all of the proportions of choice, including those for the null choices.

In the framework of the theory, the way reduction of study time produces concentering is different from the way speed instructions during test produces concentering. With speed instructions during test, $n(i, k, l)$ was unaffected during study but was reduced during the test. With reduction of study time, $n(i, k, l)$ was reduced only during study but returned to its standard state during test. Specifically, we assume that for a given condition i, $n(i, \text{study, short}) = \epsilon \cdot n(i, \text{study, long})$, where ϵ is the coefficient of reduction of feature sampling in the short-study condition, $0 < \epsilon < 1$. During the test, however, all $n(i, k, l)$ returned to their normal condition so that $n(i, \text{test, long}) = n(i, \text{test, short}) = n(i, \text{study, long})$.

The change in $n(i, k, l)$ during study had its effect in the theory on $p(i, \text{old})$, which was reduced in the short-study condition compared with the long-study condition. This was the way in which the differences between the experimental conditions were produced.

To summarize, when $n(i, k, l)$ was reduced during study by reducing study time, attention/likelihood theory again predicted the concentering of the underlying distributions (see Figure 5). The decline in recognition memory should again be observed in responses to new test words as well as old. In terms of proportions of choices, the prediction is that a short study period will produce a decline in all six choice conditions. The prediction is unique and counterintuitive in that recognition of new test words will be affected by the short- versus long-study condition, which might be assumed to affect only the old words.

In addition to the change in the main independent variable, Experiments 2a and 2b differed from Experiment 1 in the following ways. (a) The subjects carried out intentional learning without an encoding task, such as lexical decision, during the study phase. (b) The main variable, long versus short exposure of study items, was a within-subjects variable. This change was introduced to increase the power of the experiment. (c) A larger set of words was used.

Experiment 2a

Method

Each subject had four study–test sequences: two long-study and two short-study units. Each study word was presented for 1 s in short-study units and for 2 s in long-study units. The subjects did intentional learning during study. The test was self-paced forced choice.

Subjects. The subjects were 24 undergraduate students.

Procedure. An IBM-compatible personal computer and Micro Experimental Laboratory software (Schneider, 1988) were used to present stimuli and to collect response data. The subjects were given four sets of trials. In each set, a list of 60 words was presented for study, and a list of 60 test pairs was presented for the recognition test. Therefore, each subject had total of 240 study words and 240 test pairs across the four sets (60 pairs × 4 sets).

Each word in the study list was presented for either 1 or 2 s and had an interstimulus interval of 250 ms. Subjects were informed that they were going to have a recognition test.

The recognition tests were two-alternative forced choice, as in the previous experiment. During each recognition test, 60 pairs of test words were presented in the same format as in the previous experiment. There were six types of test pairs, as before: HN–LN, LO–HO, HO–HN, LO–HN, HO–LN, and LO–LN, with 10 pairs in each type. The test pairs remained on the screen until subjects responded. There was no feedback during the recognition test.

For each subject, the study–test sequence was given four times, each time with a different word list. The order of short- and long-study sequences was counterbalanced. Half of the subjects had the following study–test sequence: long study–test, short study–test, short study–test, and long study–test. The other half of the subjects had the complementary study–test sequence: short study–test, long study–test, long study–test, and short study–test.

There were 10 filler items placed at the end of each study list. The filler items were presented for 1.5 s following both the long-exposure study list (2 s) and the short-exposure study list (1 s). Subjects were not informed about the filler items and filler items were considered as part of the study list by subjects. The purpose of the filler items was to equalize response times during test. Earlier pilot work indicated a carry-over effect from study to test. Subjects in a long-study condition spent more time on test items than did subjects in a short-study condition. (See also the carry-over effect of speed versus accuracy instructions during study mentioned earlier.) Such a carry-over effect is undesirable because it makes it unclear whether an effect on proportions of choice was caused by time spent in study or time spent in test. The carryover also could be accompanied by unwanted interaction effects of response times with choice conditions. It was found that ending the study list in both conditions with a final set of filler words presented at an intermediate exposure time eliminated this carryover.

Each test list was preceded by 10 filler test pairs. The filler pairs consist of filler items from the study list and 10 new words. We did not score the filler test pairs and did not include them in the analysis.

Material. Sets of 240 high-frequency words and 240 low-frequency words were selected from Kučera and Francis (1967). High-frequency words were those that occurred 40 or more times per million words, and low-frequency words occurred 8 or fewer times per million. Word length ranged from four to nine. The two groups of words were matched on word length and concreteness value according to norms by Paivio et al. (1968). An additional 80 middle-frequency words were selected as filler items.

The 240 high-frequency words were divided into six groups with 40 words in each group. All groups were matched in terms of word frequency, concreteness, and word length. The 240 low-frequency words were also divided into six matched groups. The six groups of high-frequency words and the six groups of low-frequency words were combined to form the six groups of word pairs (HN–LN, LO–HO, HO–HN, LO–HN, HO–LN, and LO–LN), with 40 pairs in each group. Each group of 40 pairs was divided again into four sets of 10 pairs within each choice type. Across the subjects, each set of test pairs was used equally often in each of the four study–test blocks. Words in each pair were matched in word length. Words used in one condition for one subject (e.g., HN) were used in the other condition (e.g., HO) for another subject.

Results and Discussion

The main results of the experiment are summarized in the "Observed" rows in Table 3. The mirror effect was present in both short- and long-study conditions. Moreover, the short-study condition showed the predicted decline in all proportions of choice. The difference in recognition performance between long- and short-study group was statistically significant, $F(1, 23) = 22.324, p < .0005, MS_e = 0.08$. There were, as in previous results (see Experiment 1), statistically significant differences among the six choice conditions, $F(5, 115) = 39.94$, $p < .0001$, $MS_e = 0.079$, that reflected the mirror inequalities. There was no significant interaction, $F(5, 115) = 1.33$, $p > .25$, $MS_e = 0.06$.

In order to test the mirror inequalities again, we subjected the proportions of choices for four standard choice conditions to linear trend analysis. The results were statistically significant, $F(1, 115) = 39.82, p < .01, MS_e = 0.079$, with .97 of the variance accounted for. We tested the differences from .50 of proportions of choice in the two null choice conditions with t tests. The results were statistically significant. In the HN–LN condition, $t(47) = 7.55$, $p < .005$; in the LO–HO condition, $t(47) = 8.75$, $p < .005$.

We checked the response times during test to ensure that the effects found could not be ascribed to differences in response time during test. The response times are shown in Table 4. The table shows that the response times during test did not differ for long- versus short-study conditions, $F(1, 23) = 0.047$, $p > .81$, $MS_e = 0.0537$. Therefore, the performance difference could be attributed to long- versus short-study time or, in terms of the theory, to a variation of $n(i, k, l)$ during study alone. The effect of the six choice types was statistically significant, $F(5, 115) = 33.86$, $p <$

Table 3
Experiment 2a: Mean Proportions of Choice for Long and Short Study Time

	Long-study condition					
	Null pairs		Standard pairs			
M	*P*(HN, LN)	*P*(LO, HO)	*P*(HO, HN)	*P*(LO, HN)	*P*(HO, LN)	*P*(LO, LN)
Observed	.67	.65	.78	.87	.85	.90
Predicted	.68	.67	.80	.86	.86	.90
	Short-study condition					
Observed	.64	.60	.70	.76	.80	.84
Predicted	.65	.61	.72	.76	.79	.82

Note. $n = 24$. H = high-frequency words, L = low-frequency words, N = new, and O = old. Parameters: $p(\text{new}) = .09$, $N = 1{,}000$, $n(\text{high, study, long}) = 56$, $n(\text{low, study, long}) = 76$, and $\epsilon = 0.67$.

Table 4
Experiment 2a: Mean Response Times During Test for the Long- and Short-study Time Conditions

Study condition	Null pairs		Standard pairs			
	P(HN, LN)	*P*(LO, HO)	*P*(HO, HN)	*P*(LO, HN)	*P*(HO, LN)	*P*(LO, LN)
Long						
M	2,655	1,699	2,198	1,754	2,059	1,720
SD	690	374	602	404	429	502
Short						
M	2,411	1,784	2,161	1,920	2,124	1,775
SD	673	514	632	504	573	452

Note. $n = 24$. H = high-frequency words, L = low-frequency words, N = new, and O = old.

.0001, $MS_e = 0.031$. There was also a marginally significant interaction between choice type and study-time conditions, $F(5, 115) = 2.645$, $p < .03$, $MS_e = 0.0192$.

These results support a specific prediction derived from attention/likelihood theory. They also support a generalized statement that applied to all of the experiments reported here: Changes in recognition performance with new stimuli mirrored the change in recognition performance with old stimuli.

As in Experiment 1, the model was fitted to the data. The expected proportions computed from the model are shown below the observed values in the "Predicted" row in Table 3. The estimated parameters, including the parameter ϵ that reduced $n(i, k, l)$ in the short-study condition, are given at the bottom of the table. They fit the data closely. As in Experiment 1, the proportion of variance accounted for by the theory was estimated by regressing the observed on the predicted values. The proportion is $R^2 = .99$.

It should also be noted that the estimated parameters were highly similar to the corresponding parameters of Experiment 1. In both, N was fixed at 1,000. The remaining parameters for Experiment 1 and this experiment were as follows: n(high, study,.) = 53 in Experiment 1, n(high, study, long) = 56 here; n(low, study,.) = 69 in Experiment 1 and n(low, study, long) = 76 here; and p(new) = .1 in Experiment 1 and .09 here. This was expected because the materials and general procedure were highly similar in the two experiments.

Figure 7 shows the log likelihood ratio distributions of short- and long-study conditions constructed from the estimated parameters according to the six assumptions given earlier for the theory. The distributions again show the pattern depicted in Figure 5. Distances between all distributions decreased when study time was decreased.

Experiment 2b

Experiment 2b was a replication of Experiment 2a. We conducted this replication to check on the reliability of the findings of Experiment 2a. A recent study by Hirshman and Palij (1992) presented data that do not agree with ours. They carried out a sequence of comparisons of long- versus short-study times with low- and high-frequency words using a yes–no recognition test: 800 versus 1,000 ms, 800 versus 1,200 ms, and 800 versus 2,500 ms. Their concern was somewhat different from ours in that they wanted to demonstrate that increases in study time did not increase the word-frequency effect (i.e., the difference between recognition memory performance on low- and high-frequency words as measured by differences in hits and false alarms in their long- versus short-study conditions). They also used d's in their analyses. In our terms, they were concerned with demonstrating that the difference between LN and HN and the difference between LO and HO remained constant with changes

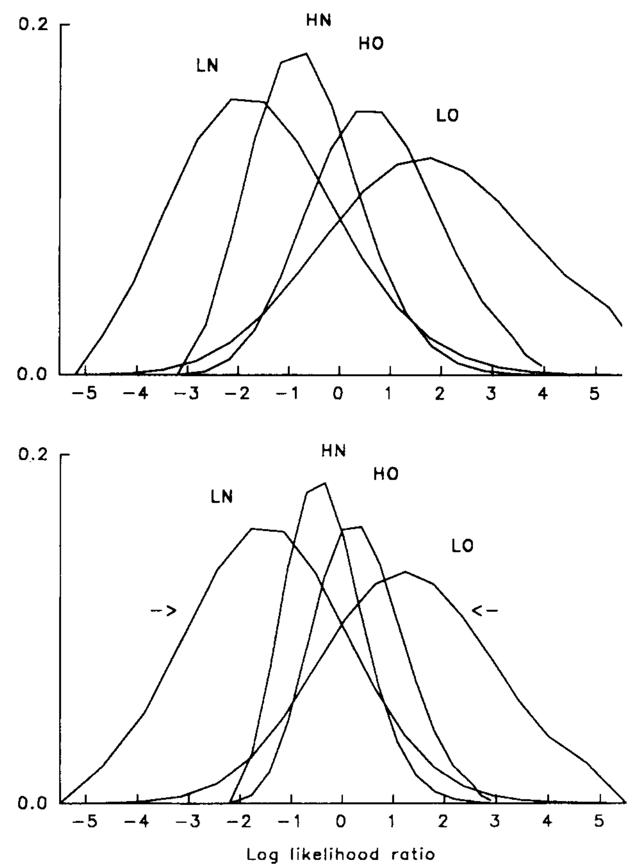

Figure 7. The log likelihood ratio distributions constructed with the parameters obtained by fitting the theory to the data of Experiment 2a. (The top panel represents the underlying distributions in the long-study condition. The bottom panel represents the distributions in the short-study condition. LO = low-frequency old, HO = high-frequency old, HN = high-frequency new, and LN = low-frequency new.)

in study time. Hirshman and Palij found that increased study time had an effect in improving performance but that neither the distance between two new distributions, LN and HN, nor the distance between the two old distributions, HO and LO, was altered. Some of their data are presented in the Appendix. Mirror effects were clear in both conditions of the experiment as well as in the other data presented in their article. Comparisons, however, of the differences in p(yes) for the new (HN vs. LN) and for the old (LO vs. HO) across the two study conditions did not show the concentering that we observed in Experiment 2a. We discuss possible reasons for the differences later. Our first concern was, however, the reliability of the findings of Experiment 2a. We therefore carried out a full replication of Experiment 2a with a new group of 24 subjects.

The results of the Experiment 2b are basically identical to the results of Experiment 2a. The main results of the experiment (proportions of choice) are summarized in the "Observed" rows in Table 5. The predicted proportions of choice based on the attention/likelihood theory are presented in the "Predicted" rows. Statistical analysis of the proportions of choice and response times gave the same results as in Experiment 2a. As expected, the estimated parameters were closely similar to corresponding parameters for Experiment 2a (see Table 3).

As in previous experiments, the proportion of variance accounted for by the theory was estimated by regressing the observed on the predicted values ($R^2 = .98$).

The difference in recognition performance between the long- and short-study conditions was statistically significant, $F(1, 23) = 11.928$, $p < .0025$, $MS_e = 0.1002$. There were, as in previous results, statistically significant differences among the six choice conditions, $F(5, 115) = 34.42$, $p < .0001$, $MS_e = 0.088$, that reflected the mirror inequalities. There was no significant interaction, $F(5, 115) = 1.46$, $p > .208$, $MS_e = 0.06$. In order to further test the mirror inequalities, we subjected the proportions of choices for the four standard choice conditions to linear trend analysis. The results of the analysis indicate that the mirror inequalities were statistically significant, $F(1, 115) = 27.65$, $p < .001$, $MS_e = 0.088$. The variance accounted for was .95. We tested the difference of the proportions of choice from .50 in the two null choice conditions with t tests. The results show that both differences were statistically significant. In the HN–LN condition, $t(47) = 7.74$, $p < .005$, and in the LO–HO condition, $t(47) = 7.06$, $p < .005$.

The response times are shown in Table 6. We were concerned with whether the response times during test would be the same in both study conditions. They were. There was no statistically significant difference, $F(1, 23) = 2.922$, $p > .1$, $MS_e = 0.041$. As in the preceding experiment, the choice conditions differed in response times, $F(5, 115) = 26.634$, $p < .001$, $MS_e = 0.025$. The interaction of study conditions with choice conditions was, however, not statistically significant, $F(5, 115) = 1.341$, $p < .25$, $MS_e = 0.0167$.

There are two possible reasons for the discrepancy between the Hirshman and Palij (1992) findings and ours. One has to do with procedure. They did not measure response times during test. We noted earlier that subjects showed carryover effects of study conditions that affected their response times during the test. If there were such carry-over effects with subjects in one condition spending a different amounts of time on low- as compared with high-frequency words in the different experimental conditions, then a pattern such as that found by Hirshman and Palij could be obtained. We discussed the problems such carry-over effects generate earlier. It was to eliminate those effects that special procedures were carried out in Experiments 2a and 2b, with intermediate exposure times used in the set of filler items at the end of the study list.

Another possible reason for the discrepancy may be in the use of the yes–no test rather than forced choice. The yes–no test is relatively insensitive, particularly to small changes in the distances between LN and HN and between LO and HO. The underlying distributions may concenter even though the yes–no data do not show the effect well. Forced-choice or confidence ratings are more sensitive and more likely to show the effect clearly.

We make this point by carrying out two steps. First, we fit attention/likelihood theory to a set of Hirshman and Palij (1992) data in the same way that we fitted it to our own. The set selected used study times (800 ms vs. 2,500 ms) that were close to ours (1,000 and 2,000 ms). An adequate fit with a theory that implies concentering would indicate that the data of a yes–no test may not fully display the changes in the underlying distributions. We then took the parameters obtained by fitting the theory to the Hirshman and Palij yes–no data and generated the corresponding predicted forced-

Table 5
Experiment 2b: Mean Proportions of Choice for Long- and Short-Study Time Conditions

	Long-study condition					
	Null pairs		Standard pairs			
M	P(HN, LN)	P(LO, HO)	P(HO, HN)	P(LO, HN)	P(HO, LN)	P(LO, LN)
Observed	.64	.68	.76	.85	.85	.89
Predicted	.66	.66	.79	.85	.85	.89
	Short-study condition					
Observed	.62	.63	.75	.79	.80	.83
Predicted	.63	.62	.73	.79	.80	.84

Note. $n = 24$. H = high-frequency words, L = low-frequency words, N = new, and O = old. Parameters: $p(\text{new}) = .086$, $N = 1{,}000$, n(high, study, long) = 53, n(low, study, long) = 72, and $\epsilon = 0.78$.

Table 6
Experiment 2b: Mean Response Times During the Test for the Long- and Short-Study Conditions

Study condition	Null pairs		Standard pairs			
	P(HN, LN)	*P*(LO, HO)	*P*(HO, HN)	*P*(LO, HN)	*P*(HO, LN)	*P*(LO, LN)
Long						
M	2,389	1,798	2,009	1,669	1,893	1,737
SD	577	599	419	397	459	482
Short						
M	2,191	1,662	1,954	1,715	1,863	1,691
SD	687	479	542	450	415	596

Note. n = 24. H = high-frequency words, L = low-frequency words, N = new, and O = old.

choice means. As we show, these predicted forced-choice means will display the concentering that is not clearly present in either the obtained or predicted yes–no data.

When we fit the theory to the data, we obtained a fit that accounted for 98% of the variance in their data. The parameters and the fit are given in the Appendix. This fit occurred despite the fact that their data did not show concentering. The theory does, however, produce the concentering in the underlying distributions that is seen neither in the actual yes–no data nor in the predicted yes–no data. That concentering can be clearly seen, however, when those same parameters are used to compute the forced-choice pattern they produce. The predicted forced-choice means for the parameters that fit their yes–no data are also given in the Appendix.

GENERAL DISCUSSION

The three studies reported here support attention/likelihood theory and its extensions by showing that they can predict the proportions of choice in the forced-choice recognition test. In the experiments, attention to items, conceptualized in the theory as $n(i, k, l)$, the amount of feature sampling, was varied in two ways: by speed versus accuracy instructions during test and by long- versus short-study time. The mirror pattern was observed under both levels of both variables. More important, both variables produced systematic, symmetric changes in the sets of underlying distributions. Variables that reduced attention either during test or study produced changes in both the distributions corresponding to new stimuli and those corresponding to old stimuli. Moreover, in conditions that reduced the attention parameter, the distributions moved toward a central point on the decision axis and also moved toward each other, showing concentering. The evidence of concentering was the consistent decrease across all proportions of choice in the disadvantageous condition.

This study showed results parallel to those of Glanzer et al. (1991). In that study, forgetting affected recognition performance on new test items as well as on old test items. In terms of the theory, the effect of delay was different from that of study time or speed versus accuracy instructions. Delay directly reduced $p(i, \text{old})$ at test. That way of reducing $p(i, \text{old})$, however, also caused concentering of underlying distributions.

The results of Experiments 2a and 2b on the effect of study time are particularly counterintuitive. Only the old stimuli were given long or short study time, and yet the recognition of new test stimuli as well as that of old test stimuli were affected.

There are three alternative approaches to explaining the mirror effect. One is a variant of attention/likelihood theory. The other two are basically different theories.

The variant of attention/likelihood theory, proposed by a reviewer, is particularly relevant to Experiments 2a and 2b.[4] In the formulation we have presented, the $p(i, \text{old})$ used in generating the predictions is the one that the subject estimates on the basis of experience in the study trial. This is the term that is inserted in Equation 2, the equation for the likelihood ratio, in all of the work reported earlier. In the alternative view, the subject does not use the estimate of $p(i, \text{old})$ based on the experience of the study trial but another estimate based solely on the experience of the test trial. We label this estimate $t(i, \text{old})$. This term $t(i, \text{old})$ and its complement replaces $p(i, \text{old})$ and its complement inside the logarithmic terms of Equation 2, effecting major changes in the likelihood ratios.

To clarify the difference, we construct two soliloquies on the part of the subject, one that corresponds to attention/likelihood theory as we have presented it here and one that corresponds to the proposed alternative. We are not asserting that such a soliloquy actually takes place. We present it solely to clarify the difference between the two formulations.

In attention/likelihood theory as presented earlier, the subject may be considered to say the following when considering an item during test: "If this were an old item like one that I studied earlier, what would it be like? If it were a new item, what would it be like?" The subject then combines those two

[4] The alternative proposed is motivated by a concern that theory places too much information in the subject. To allay this concern, Glanzer and Adams (1990, p. 13) noted that the subject's estimate of $p(i, \text{old})$ need not be accurate, that is, correspond to the actual $p(i, \text{old})$, for the mirror regularities to hold. For example, consider the case in which the subject does not estimate different $p(i, \text{old})$ for the two word classes used in the experiment. Instead of inserting $p(\text{low, old})$ and $p(\text{high, old})$ in Equation 2, when appropriate, the subject may use $p(. , \text{old})$, the mean of the two $p(i, \text{old})$. It can be shown that the main findings concerning the mirror effect still hold in this case. It can also be shown that the concentering observed in Experiments 2a and 2b also holds in this case. This variation of theory, however, still assumes that the subject's estimate is based on information from study, not solely from information during test.

estimates in a likelihood ratio (Equation 2) and uses it to make a decision.

In the proposed alternative, the subject does not estimate what an old item is expected to look like from experience in the study trial but computes it independently during the test. The subject in this soliloquy says the following: "Here is an item. I am giving it n amount of attention (drawing sample size of n items). A new item is expected to have p features marked. Therefore, I can compute, using n and p(new) and Equation 1, what an old item is expected to look like—$t(i,$ old)." The likelihood ratios are then computed from those two estimates, $t(i,$ old) and p(new), in an altered Equation 2.

The argument for this alternative is that it is simple and intuitively appealing. It does not require the subject to process information other than that available during the test trial. There are, however, two arguments against it. One is that there is no basis for asserting that the subject is restricted to information from the test situation alone. Indeed, there is ample evidence that the subject has available large amounts of information about the study trial. It has been shown that subjects remember much information specific to the study list: the presentation modality (visual vs. auditory) of study-list items, the voice (male vs. female) in which they are spoken, and the format of the letters (block vs. script). This has been shown to be true even when subjects are not instructed to attend to those characteristics or to remember them (Hintzman, Block, & Inskeep, 1972). These are not the only types of information that subjects retain about study lists. Anderson and Bower (1974) have summarized many other characteristics of study-list items that subjects remember. In addition, Anderson and Bower showed that subjects could also distinguish in their recognition responses whether the item appeared in the last study list or one before it or two before it, up to at least five study lists back. Given this capacious and discriminative memory about study lists, there is no reason to assume that the subject's memory is restricted to the items immediately available on the test list. The other and more important argument against this alternative view is that it predicts a pattern of results that is contradicted by the data. It does not produce the systematic concentering seen in Tables 3 and 5.

There are two other approaches to the mirror effect. Both of them concern only a special case—the mirror effect in word frequency—and both of them view memory strength (or familiarity) as the basis of recognition decision. The first one is presented in Bowles and Glanzer (1983) and Glanzer and Bowles (1976). In this approach, low-frequency new words are more accurately recognized as being new than high-frequency new words because high-frequency words receive more indirect marking. In other words, high-frequency new words are more likely to be responded to as old because they are more likely to be associated with old words. Low-frequency old words are better recognized than high-frequency old words because low-frequency words are assumed to have fewer meanings than high-frequency words. That makes low-frequency words easier to mark fully during study (see Glanzer & Bowles, 1976, p. 27, for details).

Gillund and Shiffrin (1984) approached the mirror effect as follows: Subjects first classify high- and low-frequency words and set different decision criteria for each. The subjects then subtract attained memory strength of a test word from each criterion and scale the value by dividing by an estimate of the standard deviation of the distractor distribution.

Both of these two approaches are restricted in that they have considered only word-frequency effects. The mirror effect can, however, be found with many other variables (Glanzer & Adams, 1985). Moreover, neither approach has been extended to explain the orderly, systematic pattern of changes in the mirror effect found in our study when study time was varied or speed versus accuracy instructions were given.

Attention/likelihood theory differs from other theories of recognition memory in that it assigns the likelihood ratio a key role in the recognition decision made for each test stimulus. The decision axis is likelihood, not as in most analyses of recognition memory, familiarity, or strength (see Kintsch, 1970, for what has become the generally accepted view). Likelihood-ratio-based recognition decisions mean that subjects jointly consider a test item's probability of being old and that of its being new when making a recognition decision. These two probabilities are compared and a decision is made on the basis of the comparison. The joint consideration produces the symmetries we have presented.

References

Anderson, J. R., & Bower, G. H. (1974). Interference in memory for multiple contexts. *Memory & Cognition, 2,* 509–514.

Bower, G. H. (1972). Stimulus-sampling theory of encoding variability. In A. W. Melton & E. Martin (Eds.), *Coding processes in human memory* (pp. 85–123). New York: Wiley.

Bowles, N., & Glanzer, M. (1983). An analysis of interference in recognition memory. *Memory & Cognition, 11,* 307–315.

Brent, R. P. (1973). *Algorithms for minimization without derivatives.* Englewood Cliffs, NJ: Prentice-Hall.

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

Estes, W. K. (1950). Toward a statistical theory of learning. *Psychological Review, 57,* 94–107.

Estes, W. K. (1955). Statistical theory of spontaneous recovery and regression. *Psychological Review, 62,* 145–154.

Forster, K. I., & Chambers, S. M. (1973). Lexical access and naming time. *Journal of Verbal Learning and Verbal Behavior, 12,* 627–635.

Gegenfurtner, K. R. (1992). PRAXIS: Brent's algorithm for function minimization. *Behavior Research Methods, Instruments, & Computers, 24,* 560–564.

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

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

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

Glanzer, M., Adams, J. K., & Iverson, G. (1991). Forgetting and the mirror effect in recognition memory: Concentering of underlying distributions. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 17,* 81–93.

Glanzer, M., & Bowles, N. (1976). Analysis of the word-frequency effect in recognition memory. *Journal of Experimental Psychology: Human Learning and Memory, 2,* 21–31.

Glanzer, M., & Ehrenreich, S. L. (1979). Structure and search of the internal lexicon. *Journal of Verbal Learning and Verbal Behavior, 18,* 381–398.

Hintzman, D. L., Block, R. A., & Inskeep, N. R. (1972). Memory for mode of input. *Journal of Verbal Learning and Verbal Behavior, 11,* 741–749.

Hirshman, E., & Palij, M. (1992). Rehearsal and the word-frequency effect in recognition memory. *Journal of Memory and Language, 31,* 477–484.

Kintsch, W. (1970). *Learning, memory, and conceptual processes.* New York: Wiley.

Kučera, H., & Francis, W. N. (1967). *Computational analysis of present-day American English.* Providence, RI: Brown University Press.

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

Scarborough, D. L., Cortese, C., & Scarborough, H. S. (1977). Frequency and repetition effects in lexical memory. *Journal of Experimental Psychology: Human Perception and Performance, 3,* 1–17.

Schneider, W. (1988). Micro Experimental Laboratory: An integrated system for IBM PC compatibles. *Behavior Research Methods, Instruments, & Computers, 20,* 206–217.

Appendix

The Observed and Predicted yes–no Proportions for Hirshman and Palij's (1992) Experiment 3

For the long-study condition (2,500 ms):

	LN	HN	HO	LO
Observed	.08	.15	.64	.72
Predicted	.11	.14	.61	.75.

For the short-study condition (800 ms):

	LN	HN	HO	LO
Observed	.17	.25	.48	.64
Predicted	.18	.24	.52	.58.

L = low-frequency words, H = high-frequency words, N = new, and O = old. The parameters used were the following: N = 1,000 (preset), p(new) = .1188, n(high, study, long) = 65, n(low, study, long) = 83, ϵ = .57, Criterion c = .2401 (because the data show strong bias), R^2 = .98.

Note the unclarity of the concentering in the predicted means. In particular, note that the predicted means show a reversal with respect to concentering in the difference between LN and HN in the long (.03) versus the short (.06) conditions. That difference was smaller in the 2,500-ms condition than in the 800-ms condition instead of being larger.

The predicted forced choices for the same parameters were the following.

	P(HN, LN)	P(LO, HO)	P(HO, HN)	P(LO, HN)	P(HO, LN)	P(LO, LN)
Long	.66	.65	.82	.87	.87	.90
Short	.60	.59	.70	.74	.75	.78.

Note the completely consistent drops in the proportions that are the sign of concentering. The forced-choice predictions for the same parameters show the underlying concentering that yes–no predictions and data do not.

Received September 9, 1991
Revision received August 4, 1992
Accepted August 14, 1992 ■