The Mirror Effect and Attention–Likelihood Theory: A Reflective Analysis

Bennet B. Murdock
University of Toronto

The mirror effect refers to findings from studies of recognition memory consistent with the idea that the underlying "strength" distributions are symmetric around their midpoint separating studied and nonstudied items. Attention–likelihood theory assumes underlying binomial distributions of marked features and claims that old-item differences result from differential attention across conditions during study. The symmetry arises because subjects use the likelihood ratio as the basis for decision. The author analyzes the model and argues that one of the main criticisms (the complexity of the likelihood-ratio decision rule) is unwarranted. A further analysis shows that other distributions (the Poisson and the hypergeometric) can also produce a mirror effect. Even with the binomial distribution, a variety of parameter values can produce a mirror effect, and with the right combination of parameter values, differential attention across conditions is not necessary for a mirror effect to occur.

The mirror effect refers to the fact that in a study–test procedure, experimental variables, such as word frequency, that result in greater recognition accuracy for old (studied) items also result in greater recognition accuracy for new (nonstudied) items. Thus, if the hit rate for low-frequency old items is higher than the hit rate for high-frequency old items, then the false-alarm rate for low-frequency new items will also be lower than the false-alarm rate for high-frequency new items. This result would arise if the underlying strength or familiarity distributions were as shown in Figure 1, where in this case AN would be low-frequency new items, BN would be high-frequency new items, BO would be high-frequency old items, and AO would be low-frequency old items.

The mirror effect has been intensively studied by Glanzer and his colleagues (e.g., Glanzer & Adams, 1985; Glanzer, Adams, & Iverson, 1991; Glanzer, Adams, Iverson, & Kim, 1993; Kim & Glanzer, 1993), and the present description comes from their analysis. The effect is important because it is counter to signal-detection accounts of recognition memory and to global-matching memory models such as SAM (search of associative memory; Gillund & Shiffrin, 1984), MINERVA (Hintzman, 1988), CHARM (composite holographic associative recall model; Metcalfe-Eich, 1982), TODAM (theory of distributed associative memory; Murdock, 1982), and the matrix model (Humphreys, Pike, Bain, & Tehan, 1989). What is problematic for signal-detection and global-matching models is that the AN distribution is below the BN distribution; there is no reason for new A and B items to differ because neither have been seen before in the experiment.

The most convincing evidence for the mirror effect comes from the forced-choice inequalities. With four conditions (AN, BN, BO, and AO), there are $\binom{4}{2}$ or six pairwise comparisons. Following Glanzer et al. (1993), if $P(\text{BO,BN})$ denotes the proportion of correct responses in a two-alternative forced-choice situation that pits a BO item against a BN item, then the implication of Figure 1 is that $P(\text{BO,BN}) < P(\text{BO,AN}) = P(\text{AO,BN}) < P(\text{AO,AN})$ and that, for null choices (i.e., comparing items from the two new-item distributions or items from the two old-item distributions), $P(\text{BN,AN}) = P(\text{AO,BO}) > 0.5$. For null choices, a "correct" response is arbitrarily defined as the B item for BN,AN but the A item for AO,BO. These results seem to be typical of the data (e.g., Glanzer et al., 1993).

Attention–Likelihood Theory

Attention–likelihood theory (ALT) is an attempt by Glanzer and his colleagues (e.g., Glanzer & Adams, 1985; Glanzer et al., 1991; Glanzer et al., 1993; Kim & Glanzer, 1993) to explain the mirror effect in general and these forced-choice inequalities in particular. ALT says that the mean difference between the AO and the BO distributions comes about because during study subjects pay more attention to (or mark or tag more features from) the items in Condition A than in Condition B, but the placement of the four distributions along the decision axis comes about because subjects transform strength or familiarity to log likelihood ratios and use the log likelihood ratio, not strength, as the basis for their decision. (The distributions shown in Figure 1 are plotted in terms of log likelihood ratio, not strength.)

More particularly, ALT assumes that the underlying strength distributions are binomial. A binomial distribution characterizes the probability of x successes in a sample of
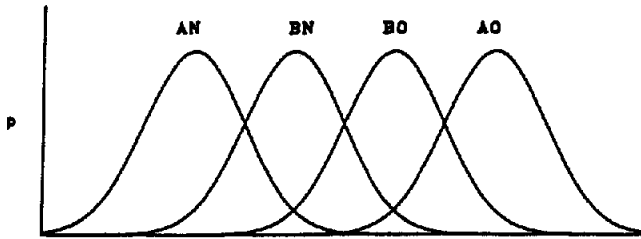
Figure 1. Hypothetical strength distributions for low-frequency new items (AN), high-frequency new items (BN), high-frequency old items (BO), and low-frequency old items (AO). p = probability. From "The Regularities of Recognition Memory," by M. Glanzer, J. K. Adams, G. J. Iverson, and K. Kim, 1993, *Psychological Review, 100*, p. 547. Copyright 1993 by the American Psychological Association. Reprinted with permission of the author.

size n from a population where the probability of obtaining a success is p. If x is the independent variable, then a binomial distribution in x is a function of the two parameters n and p. In a coin-tossing situation, for instance, if x were the number of heads, n would be the number of coins, p would be the probability of a head, and if $Bi(x; n, p)$ is a binomial in x with parameters n and p, then

$$Bi(x; n, p) = \binom{n}{x} p^x (1 - p)^{n-x}.$$

The probability of all possible outcomes $0 \le x \le n$ is given by this distribution. In the coin-toss example $p = .50$, but it need not be.

ALT is a stimulus-sampling theory like the stimulus-sampling theory of encoding variability of Bower (1972), after Estes (1959). Items are represented as lists of features, and each feature is either marked or tagged or else it is not marked or tagged. According to ALT, subjects pay more attention to the items in Condition A than in Condition B, so $n(A)$, the number of features marked in Condition A, is greater than $n(B)$, the number of features marked in Condition B. Because the probability that any particular feature will be marked is the number of features marked divided by the total number of features (N), $p(i, \text{old})$, the proportion of marked features in the old-item distribution of Condition i is proportional to $n(i)/N$, where $n(i)$ is the number of marked features in Condition i, i being A or B. Call this ratio $\alpha(i)$. It is assumed that new items will have some proportion of features already marked; call this $p(\text{new})$. For the old-item distribution of Condition i, when a study item is presented, features that are already marked or tagged will stay marked, but unmarked features are marked with probability $\alpha(i)$. Therefore,

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

Now $p(\text{new})$ is the same for Conditions A and B but, because $n(A) > n(B)$, $p(A, \text{old}) > p(B, \text{old})$.

As Glanzer et al. (1993) said, if subjects made their decision in terms of the number of marked features, ALT would be a standard strength theory. Instead, subjects are assumed to base their decision on the log likelihood ratio, where the log likelihood ratio is a function of x. The likelihood ratio is the probability of a particular x value from one distribution divided by the probability of the same x value from another distribution, and the log likelihood ratio is the logarithm of this ratio.

To illustrate with an example, assume that $n(A) = 3$, $n(B) = 2$, $p(A, \text{old}) = 0.4$, $p(B, \text{old}) = 0.2$, and $p(\text{new}) = 0.1$. Four binomial distributions are needed in x (x is the number of marked features), and call them $f_1(x), f_2(x), f_3(x)$, and $f_4(x)$, where

$$f_1(x) = Bi[x; n(A), p(\text{new})] \text{ for AN,}$$

$$f_2(x) = Bi[x; n(B), p(\text{new})] \text{ for BN,}$$

$$f_3(x) = Bi[x; n(B), p(B, \text{old})] \text{ for BO,}$$

and

$$f_4(x) = Bi[x; n(A), p(A, \text{old})] \text{ for AO.}$$

These are the marking distributions for new and old items in Conditions A and B with parameters $n(A)$, $n(B)$, $p(\text{new})$, $p(A, \text{old})$, and $p(B, \text{old})$. These four distributions are shown in Table 1.

According to ALT, the log likelihood ratio distributions for Conditions A and B, not the marking distributions, are the basis for decision. The log likelihood ratio at x for Condition A, which is symbolized as $\ln [\lambda_A(x)]$, is $\ln [f_4(x)/f_1(x)]$. For Condition B, $\ln [\lambda_B(x)]$ is $\ln [f_3(x)/f_2(x)]$. These log likelihood ratios, $\ln [\lambda_A(x)]$ and $\ln [\lambda_B(x)]$, are simply the natural logarithms (log base e or Napierian logarithms denoted as ln) of the ratios of the two ordinates (probability values), where in both cases (Condition A and Condition B) the ratio is old/new. These log likelihood ratios are functions of x because both $\ln [\lambda_A(x)]$ and $\ln [\lambda_B(x)]$ vary with x. To simplify the notation, call the log likelihood ratios $g_A(x)$ and $g_B(x)$, where

$$g_A(x) = \ln [\lambda_A(x)] = \ln \frac{f_4(x)}{f_1(x)},$$

Table 1
Binomial Distributions for Conditions AN, BN, BO, and AO

x	AN $f_1(x)$	BN $f_2(x)$	BO $f_3(x)$	AO $f_4(x)$
0	.729	.810	.640	.216
1	.243	.180	.320	.432
2	.027	.010	.040	.288
3	.001			.064

Note. AN = low-frequency new items; BN = high-frequency new items; BO = high-frequency old items; AO = low-frequency old items; x = the number of marked features.

and

$$g_B(x) = \ln [\lambda_B(x)] = \ln \frac{f_3(x)}{f_2(x)}.$$

The numerical values of these log-likelihood ratios $g_A(x)$ and $g_B(x)$ are shown in Table 2.

If the forced-choice comparisons were always between an old A item and a new A item or between an old B item and a new B item, then all that is needed would be $g_j(x)$, where j = A in the first case and j = B in the second case. However, there are six possibilities (AO/AN, BO/BN, AO/BN, BO/AN, AO/BO, and BN/AN), so two log likelihood ratios are needed, depending on the particular forced-choice comparison. Let $g_j(x)$ be the log likelihood ratio for the stronger item, as defined by the four distributions shown in Figure 1, and let $g_k(x)$ be the log likelihood ratio for the weaker item, as defined similarly. Then j and k will depend on the particular forced-choice comparison. Thus, for any forced-choice comparison, j and k can be determined and then the appropriate log likelihood ratios can be found.

Now a second set of binomial distributions must be constructed (Glanzer et al., 1993, footnote 4) that can be symbolized Bi'_j and Bi'_k. These are binomial distributions not in x but in $g_j(x)$ and $g_k(x)$; that is, Bi'_j is a binomial distribution in $g_j(x)$—$g_j(x)$ is the independent variable—and Bi'_k is a binomial distribution in $g_k(x)$—$g_k(x)$ is the independent variable. The decision axis has been transformed or rescaled from features (x) to log likelihood ratios, $g_j(x)$ and $g_k(x)$. The probabilities of this second set of binomial distributions are those associated with the parent binomials $f_i(x)$, where i = 1, 2, 3, or 4, depending on the type of pair being tested.

Note that these rescaled binomial distributions are constructed binomials. They are constructed by taking particular probability values from the parent binomial distributions and by associating them with the appropriate log likelihood ratios. The log likelihood ratio then functions as the decision axis in the forced-choice situation. Even if the conditions (A or B) are the same and the gamma function is used for the factorials, the binomial probability of a likelihood ratio cannot be computed because the right result will not be obtained, but the binomial probability values can be associated with the corresponding likelihood ratios.

As a specific example, suppose that the forced-choice comparison is an AO item compared with a BN item. Then j = A and k = B, and the parent distributions are $f_4(x)$ and

Table 2
Log Likelihood Ratios $g_A(x)$ and $g_B(x)$ for Conditions A and B, Respectively

x	$g_A(x)$	$g_B(x)$
0	−1.216	−0.236
1	0.575	0.575
2	2.367	1.386
3	4.159	

Note. x = number of marked features.

Table 3
Transformed (Rescaled) Binomial Distributions Bi'_j (for AO) Compared With Bi'_k (for BN)

Bi'_j		Bi'_k	
$g_A(x)$	$f_4(x)$	$g_B(x)$	$f_2(x)$
−1.216	.216	−0.236	.810
0.575	.432	0.575	.180
2.367	.288	1.386	.010
4.159	.064		

Note. AO = low-frequency old items; BN = high-frequency new item; $g_A(x)$ and $g_B(x)$ = log likelihood ratios for Conditions A and B, respectively; $f_4(x)$ and $f_2(x)$ = binomial distributions for Conditions AO and BN.

$f_2(x)$, respectively. The transformed binomial distributions are shown in Table 3. The independent variables are the likelihood ratios, and the associated probabilities are those of the parent distributions. This transformation works for any forced-choice comparison. For any comparison, determine j and k, find $g_j(x)$ and $g_k(x)$ (the log likelihood ratios) from Table 2, and then select the parent distributions $f_j(x)$ and $f_k(x)$ from Table 1. Then have the transformed or rescaled binomial distributions Bi'_j and Bi'_k form the basis for decision.

The standard measure of performance in a forced-choice recognition experiment is the probability of a correct response (PC), not d'. If j is the stronger binomial and k is the weaker binomial condition, then, for discrete distributions, PC_{jk} is the probability that, for each log likelihood value of binomial j, the log likelihood value of binomial k is below that of binomial j summed over all values of binomial j (plus .5 times the joint probability given a tie). Thus,

$$PC_{jk} = \sum_u Bi'_j(u) \left[\sum_{v<u} Bi'_k(v) + .5 \, Bi'_k(v) \right], \qquad (2)$$

for the u and v log likelihood values of the j- and kth binomials where, as always, j and k are given by the particular forced-choice comparison.

Table 3 shows $g_A(x)$ and $f_4(x)$ for Bi'_j (AO) and $g_B(x)$ and $f_2(x)$ for Bi'_k (BN). Because the log likelihood ratio is the decision axis, $g_A(x)$ is the rescaled x values for Bi'_j, $g_B(x)$ is the rescaled x values for Bi'_k, and $f_4(x)$ and $f_2(x)$ are the associated probability values for AO and BN, respectively. Comparison of Table 3 with Tables 1 and 2 will show how these rescaled or transformed binomial distributions are constructed.

Table 4 shows how Equation 2 is computed. Set up a table with the two row headings u (i.e., $g_A(x)$) and $Bi'_j(u)$ and the two column headings v (i.e., $g_B(x)$) and $Bi'_k(v)$. The cell entries are given for all cases that satisfy the inequality $(v < u)$. They are the joint probabilities of $Bi'_j(u)$ and $Bi'_k(v)$. Then PC_{jk} is the weighted sum of all the joint probabilities, where ties are weighted by .5 (for guessing). For this example, PC_{jk} = .742.

The final result (PC_{jk}) depends on the rescaling or change of variable (i.e., the change from features to log likelihood ratios). The ordinate (probability value) is unaffected by this

Table 4
Table to Illustrate the Computation of Equation 2

u	$Bi'_j(u)$	v: -0.236 $Bi'_k(v)$: .810	0.575 .180	1.386 .010
-1.216	.216			
0.575	.432	.350	.5 (.078)	
2.367	.288	.233	.052	.003
4.159	.064	.052	.012	.001

Note. The sum of the table entires (.742) is PC_{jk}, the weighted sum of all the joint probabilities.

change of variable, but the abscissa (the basis for decision) is changed in a significant way. To illustrate with the parameter values suggested by Glanzer et al. (1993; $N = 1,000$, $n(A) = 60$, $n(B) = 40$, and $p(\text{new}) = .10$), the binomial distributions in x (the untransformed abscissa) are shown in Figure 2 and the binomial distributions in ln λ (the transformed abscissa) are shown in Figure 3. Unlike Figure 2, the distributions in ln λ in Figure 3 are lined up very much like those of Figure 1. The AN distribution is below the BN distribution, whereas the AO distribution is above the BO distribution.

The distributions in Figure 3 are somewhat different in shape from the distributions in Figure 1, and one might wonder what Figure 3 has to do with the mirror effect. If the underlying distributions were as shown in Figure 1, mirror effect data would be obtained, but mirror effect data do not require Figure 1 distributions. Other underlying distributions can generate good approximations to mirror effect data. Glanzer et al. (1993) used binomial distributions with these parameters to illustrate the workings of ALT, so perhaps the

distributions in Figure 3, not Figure 1, should be considered the prototypical underlying distributions for the mirror effect.

This transformation (or rescaling) from features (x) to log likelihood ratios is the key to ALT; it is what enables ALT to explain the mirror effect. With this transformation, one can explain the yes–no data, mean confidence ratings, and the forced-choice inequalities very nicely (see Glanzer et al., 1993, for details). In the remainder of the article, I analyze ALT more critically.

Analysis

Overall, ALT is a simple and elegant model that explains some puzzling data. It provides an explicit account of the decision rule used by subjects, a notable weakness of global matching models. It takes us beyond strength theory in understanding recognition memory. However, it has been criticized.

Criticisms

ALT is often criticized as being complex; how can it be called simple? As Glanzer et al. (1993) have noted, subjects need only to know the number of marked features in a test item and the class (A or B) to which it belongs. This is enough to compute the log likelihood ratio that serves as the basis for decision. The mirror effect will result even if their knowledge is approximate rather than exact (Glanzer et al., 1993).

It is important to understand that the subject need not go through all the calculations detailed above. The subject must
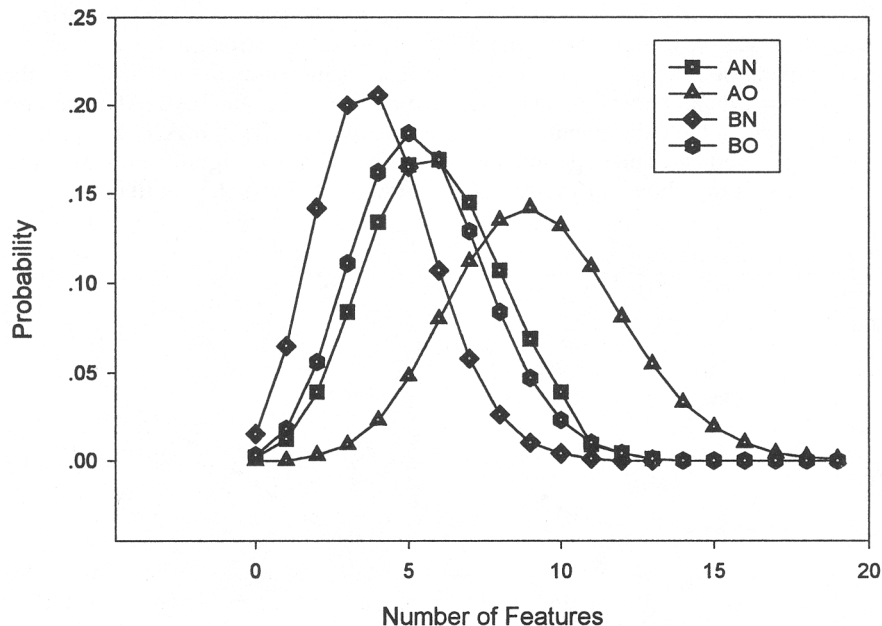

Figure 2. Binomial strength distributions for AN, BN, BO, and AO. AN = low-frequency new items; BN = high-frequency new items; BO = high-frequency old items; AO = low-frequency old items.
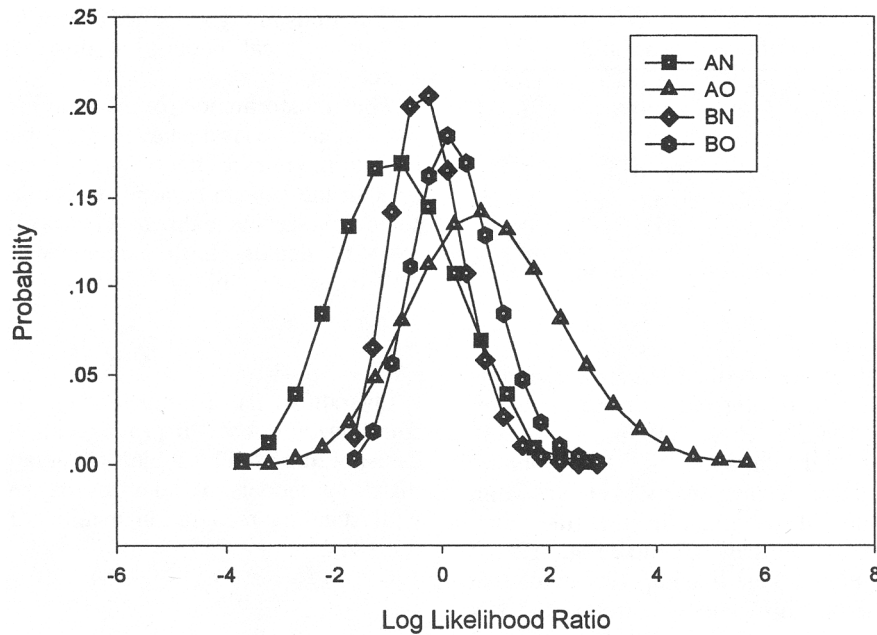
Figure 3. Transformed binomial distributions rescaled to the log likelihood ratios. AN = low-frequency new items; BN = high-frequency new items; BO = high-frequency old items; AO = low-frequency old items.

compute a log likelihood ratio but not the associated probabilities (i.e., $f_4(x)$ and $f_2(x)$ in Table 3). The experimenter must know these probabilities to derive predictions from the model, but that is quite a different matter. In fact, the subject cannot know these probabilities because their knowledge requires knowing whether each test item is old or new; that is, the subject would have to know the parent distribution—$f_i(x)$ in Table 1—for each test item. But if the subject knew whether each item was old or new, there would be no need to compute a log likelihood ratio.

Researchers must distinguish, then, between what the subject must do to respond and what the experimenter must do to determine what the model predicts. The experimenter must indeed go through the calculations shown in Tables 1–4

to determine the six forced-choice results predicted by any given set of parameter values, but the subject needs only to compute or guesstimate a log likelihood ratio.

A likelihood ratio is simply the odds for one outcome relative to another outcome and, one way or another, people probably do this many times each day when subjective probabilities are computed. A logarithmic transformation (for log likelihood ratios) linearizes the function relating likelihood to strength (i.e., number of marked features). Also, it increases the sensitivity of the decision axis. As an example, for "standard" ALT parameters for Condition A, that is, $n(AO) = n(AN) = 60$, $p(A, old) = 0.154$, and $p(new) = 0.10$, Figure 4A shows the likelihood ratio and Figure 4B shows the log likelihood ratio as a function of the
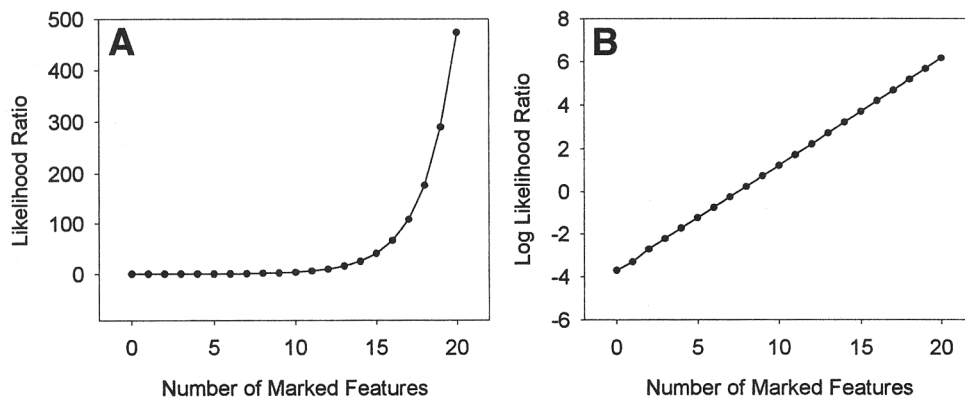

Figure 4. (A) Likelihood ratio and (B) log likelihood ratio as a function of the number of marked features.

number of marked features over the range of 0–20. Clearly, the log likelihood ratio is a much better basis for decision than the likelihood ratio.[1]

Computing a log likelihood ratio may seem complex, but consider another example. It seems to be accepted without question that a neural network in a connectionist model can compute a logistic activation function (no relation to a log likelihood ratio). Without parameters, a logistic activation function $f(x)$ is

$$f(x) = \frac{1}{1 + e^{-x}}.$$

If a neural network can compute this, it does not seem unreasonable to suggest that people can also compute the log likelihood ratios associated with various strength values.

If it is allowed that people can compute log likelihood ratios (or simply LLRs), how do subjects in an experimental session do this selectively for A and B items? Although they see A and B items in the study phase, they do not encounter the new A and B items until the test phase, so how do they learn the LLRs for Conditions A and B? When shown A and B items at study, do they conjure up comparable A and B new items with the same number of marked features? Somehow this seems unlikely, and ALT has not addressed this issue. However, as one reviewer pointed out, the subject has experience with new items at two points: during the learning and test phases. The new items encountered early in the test phase may reintegrate the encounters during the learning phase (where, of course, all items were new).[2]

If it is argued that subjects quickly learn the LLRs during the first few test trials, then one could well ask how they do this. It presupposes the ability to discriminate old and new items, but this brings up the circularity problem noted above; namely, if they can discriminate old and new items, why do they need LLRs? A problem for this learning argument is that study–test experiments almost always find the best performance at the beginning of the test phase (e.g., Murdock & Anderson, 1975, Figure 4). This result would seem to speak against the learning hypothesis, although it could be that intratest interference effects are overriding the learning effect.

A further problem comes from a study by Hockley, Hemsworth, and Consoli (in press). Generally, studies of the mirror effect present both A and B items during the study phase. Using a sunglasses manipulation, they presented lists of faces (the A items) without sunglasses during study and tested both old and new faces with and without sunglasses. (Faces with sunglasses were the B items.) They found a very clear mirror effect; the proportion of "old" responses was higher for new faces with sunglasses than for new faces without sunglasses (.29 to .20), but the proportions of "old" responses was lower for old faces with sunglasses than for old faces without sunglasses (.78 to .87). This carries the problem one level deeper; not only must subjects be able to learn LLRs during study when no new items are presented, but apparently they must also be able to learn differential

LLRs when only one class of items is presented during study.[3]

As the same reviewer pointed out, the Hockley et al. (1996) results suggest that the sunglasses manipulation may have affected the sample sizes $n(A)$ and $n(B)$ at test only. On the recognition test, we can assume equal learning by letting $p(A, \text{old}) = p(B, \text{old})$, and can also assume as usual that $p(A, \text{new}) = p(B, \text{new}) = p(\text{new})$. If it can be assumed that the sunglasses condition elicits a change in the number of features sampled at test, specifically $n(B) < n(A)$, the mirror effect should still hold. That is, despite equal marking of AO and BO (and of course AN and BN), the reduced sampling for the B condition should produce likelihood ratio distributions that lie inside of the corresponding distributions for the A condition. The notion of reduced $n(i)$ for the sunglasses condition, given that the eyes bear substantial information, seems intuitively appealing.

Hintzman, Caulton, and Curran (1994) have shown the existence of a mirror effect at the shortest lags in a response-signal method and have said that "this approach [ALT] requires that the subject know, or be able to quickly estimate, the entirety of each old and new distribution at test and so implies an even bigger processing burden than does Z-score rescaling" (pp. 287–288). However, Glanzer et al. (1993) have argued that all subjects need to know is the number of marked features, the stimulus class (i.e., A or B), and the LLRs, and my analysis would seem to support this view. Also, presumably it does not take much time for neural networks to compute activation functions, and the same may be true for LLRs.

Other Distributions

Much of the criticism of ALT seems to have focused on the use of the LLR as the basis for decision, but this criticism may be misplaced. Other distributions can generate a mirror effect, and there is even some precedent for using other distributions. Murdock (1970) used a likelihood-ratio model and compared underlying binomial, Poisson, and geometric distributions. Because Poisson and geometric distributions are one-parameter distributions, one could use either of them in ALT if one assumed the parameter (μ in the Poisson or p in the geometric) differed in Condition A and Condition B and for old and new items.

Good forced-choice results can be obtained with underlying Poisson distributions. For instance, with $\mu(\text{new}) = 1.0$, $\mu(B, \text{old}) = 1.5$, and $\mu(A, \text{old}) = 2.0$, the forced-choice inequalities are 0.618, 0.693, 0.668, and 0.711 for BO/BN, BO/AN, AO/BN, and AO,AN, respectively. The null choices are 0.590 and 0.616 for AO/BO and BN/AN, respectively. Of course, this is not surprising because the Poisson is a

---

[1] The log likelihood ratio for any pair of binomials will always be a linear function as long as they have the same value of n.

[2] I would like to thank John Adams for this suggestion.

[3] This result would seem to fit well with the differential guessing bias explanation suggested by Greene (1996). However, Glanzer, Kim, & Adams (in press) have recently presented some evidence against Greene's explanation.
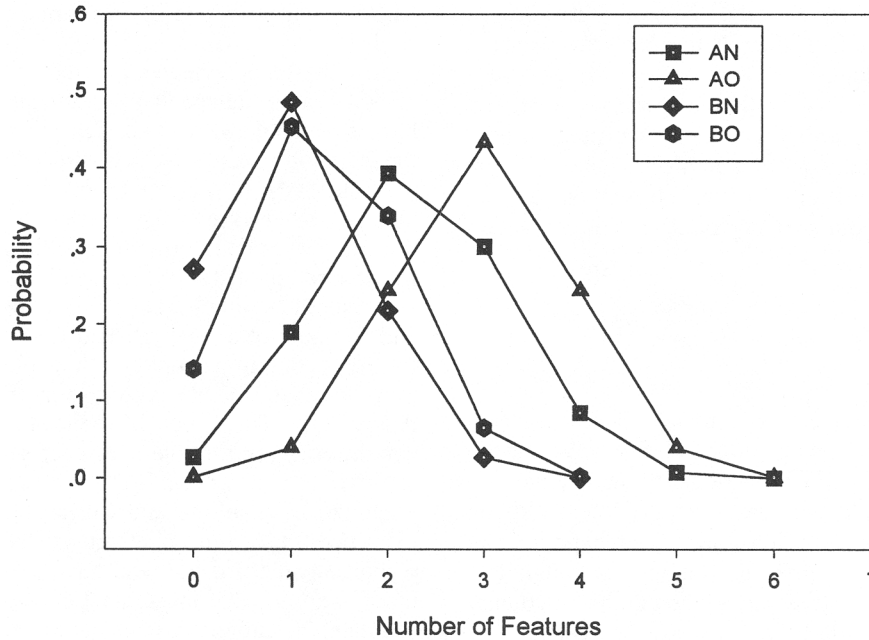
Figure 5. Hypergeometric strength distributions for AN, BN, BO, and AO. AN = low-frequency new items; BN = high-frequency new items; BO = high-frequency old items; AO = low-frequency old items.

good approximation to the binomial with small values of p and large values of n. However, I have not been able to find parameter values that give good forced-choice results if one assumes underlying geometric distributions.

Another distribution that can give a mirror effect is the hypergeometric distribution. The hypergeometric distribution is the distribution when one is sampling without replacement from a population of finite size. (By way of contrast, the binomial distribution is the distribution when one is sampling with replacement from a population of infinite size.) For a population of n elements of which $n1$ are "successes" (targets, marked element, or whatever) and $n2$ are failures, if r elements are chosen randomly, sampling without replacement, then the probability $q(k)$ of k successes is

$$q(k) = \frac{\binom{n1}{k}\binom{r - n1}{r - k}}{\binom{n}{r}},$$

for $k = 0, 1, 2, \ldots, \min(n1, r)$ (Feller, 1968). As an example that gives a mirror effect, for $n(\text{new}) = 16$, $n(\text{old}) = 12$, $n1(A) = r(A) = 6$, and $n1(B) = r(B) = 4$, the predicted forced-choice results are .61, .67, .68, .70, .57, and .57 for BO/BN, BO/AN, AO/BN, AO/AN, AO/BO, and BN/AN, respectively.[4]

What psychological interpretation could this be given? Here is one possibility. A items being more memorable (distinctive) have more marked features than B items.

However, A and B items have the same total number of features. Study reduces the number of unmarked features but equally for old and new items. Assume the sample size (r) is the same as the number of marked features so $r(A) = n(A) > r(B) = n(B)$ for both old and new items. Thus, in the present example, $r(A, \text{new}) = r(A, \text{old}) = n1(A, \text{new}) = n1(A, \text{old}) = 6 > r(B, \text{new}) = r(B, \text{old}) = n1(B, \text{new}) = n1(A, \text{old}) = 4$, and $n(A, \text{new}) = n(B, \text{new}) = 16 > n(A, \text{old}) = n(B, \text{old}) = 12$. If subjects used the LLR as the basis for their decision, then a mirror effect would result.

If one computes the probability that a given feature is marked ($p = n1/n$), then $p(A, \text{old}) > P(A, \text{new}) > p(B, \text{old}) > p(B, \text{new})$, but in terms of means (the mean of a hypergeometric distribution is $r*n1/n$) $\mu(A, \text{old}) > \mu(A, \text{new}) > \mu(B, \text{old}) > \mu(B, \text{new})$. Thus, for the means (of the hypergeometric distribution), although the old-item means are greater than the new-item means for both A and B items, unlike ALT, the new-item mean for A items is above both the B-new item mean and the B-old item mean.

The feature distributions for the hypergeometric distributions are shown in Figure 5. As can be seen, both the means and the variances of the A items are greater than the means and the variances of the B items. The rescaled log likelihood distributions are plotted in Figure 6; the means are now appropriately ordered, but the A-item variances are still larger than the B-item variances.

The means and variances of the log likelihood distribu-

---

[4] A sample program illustrating this computation written in Mathematica 3.0 is available on my home page at http://www.psych.utoronto.ca.
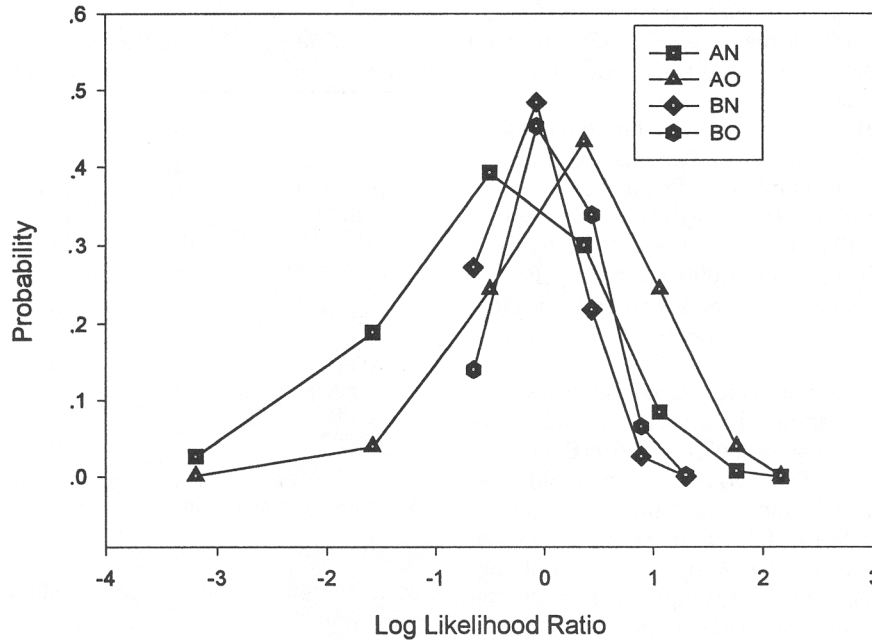
Figure 6. Transformed hypergeometric distributions rescaled to the log likelihood ratios. AN = low-frequency new items; BN = high-frequency new items; BO = high-frequency old items; AO = low-frequency old items.

tions for both the hypergeometric and the binomial distributions are shown in Table 5. The same pattern occurs for both means and variances. For both hypergeometric and binomial distributions, the means are roughly symmetrical around zero, and the A item variances for both old and new items are appreciably larger than the B item variances. These characteristics are probably the essence not only of the mirror effect but also of the other two regularities as well (the slope regularity and concentering; see Glanzer et al., 1993), so the hypergeometric distribution, as well as the binomial, seems to have all the requisite characteristics to fit the data.

The hypergeometric distribution can be negatively skewed as well as positively skewed, whereas the binomial distribution can only be positively skewed if $p < .5$. Although some accounts of the hypergeometric distribution emphasize its similarity to the binomial distribution, the two distributions

clearly can differ. In the present case, with the binomial distribution, the presentation of an item increases the number of marked features, whereas, with the hypergeometric distribution, the presentation of an item decreases the number of unmarked features.

Another difference is sampling with replacement (binomial) or sampling without replacement (hypergeometric). Sampling with replacement implies a serial process (items are sampled one by one and a running count is kept of the number of marked features), whereas sampling without replacement implies a parallel process (e.g., the complete sample is taken and then the marked features are counted). The data from a speed–accuracy trade-off study for item recognition by McElree and Dosher (1993) was interpreted as more consistent with a parallel process than with a serial process.

I am not claiming that the hypergeometric model is preferable to the binomial model of ALT. The main point here is simply to demonstrate that a rather different set of assumptions can also generate a mirror effect. This means that binomial distributions are not essential; Poisson and hypergeometric distributions can also do the job. This is an indication that perhaps other alternatives should be considered and tested against the binomial distribution.

Other Parameter Values

Not only can other distributions generate a mirror effect but the binomial distribution can also generate a mirror effect with a variety of parameter values. For instance, with $n(A) > n(B)$, a mirror effect can be obtained with various

Table 5
*Means and Variances for the Four Conditions
for the Hypergeometric and the Binomial Distributions*

Condition	Hypergeometric		Binomial	
	M	Variance	M	Variance
AN	−0.358	0.855	−0.751	1.316
BN	−0.089	0.176	−0.240	0.437
BO	0.089	0.177	0.262	0.570
AO	0.302	0.516	0.849	1.905

Note. AN = low-frequency new items; BN = high-frequency new items; BO = high-frequency old items; AO = low-frequency old items. The binomial parameters are $n(A) = 60$, $n(B) = 40$, $p(new) = .10$, $p(A, old) = .154$, and $p(B, old) = .136$.

combinations of $p(A, old)$, $p(B, old)$, and $p(new)$. On the other hand, if $n(A) = n(B)$, it may not be possible to get a mirror effect with the same combination of $p(A, old)$, $p(B, old)$, and $p(new)$ values.

There are actually six parameters for the binomial model: $N(A)$ and $N(B)$, the total number of features for A and B items; $n(A)$ and $n(B)$, the number of features sampled or attended to at study, and $p(AN)$ and $p(BN)$, the probability that any given feature will be marked for new A and B items. ALT assumes that $N(A) = N(B) = 1,000$, $p(AN) = p(BN) = .10$. It obtains $p(i, old)$ using $\alpha(i) = n(i)/N$ and assumes that subjects pay more attention to A items so $p(A, old) > p(B, old)$.

To illustrate that various parameter values can generate a mirror effect, the typical parameter values for ALT (Glanzer et al., 1993) are $p(A, old) = .154$, $p(B, old) = .136$, $p(B, new) = p(A, new) = p(new) = .10$. Thus, $p(A, old) > p(B, old)$, and the predicted mirror effect for the six forced-choice inequalities are shown as Example 1 in Table 6. In Example 2 I show the predicted mirror effect for $p(A, old) = p(B, old)$ but $p(B,new) > p(A,new)$, and in Example 3 I show the predicted mirror effect with $p(A, old) = p(B, old) > p(B, new) = p(A, new)$. In Example 4 I show the predicted mirror effect for $p(A, old) > p(B, old) > p(B, new) > p(A, new)$. All four examples show a predicted mirror effect despite the variation in the probability values, and the fourth example is the cleanest of all.[5]

On the other hand, in Table 7 I show the comparable results when $n(AO) = n(AN) = n(BO) = n(BN)$. In no case is there a reasonable mirror effect. I would not go so far as to claim that $n(A) > n(B)$ is a necessary condition for ALT to predict a mirror effect but, at least for these four examples, all cases that show a mirror effect with $n(A) > n(B)$ in Table 6 fail to show a mirror effect with $n(A) = n(B)$ in Table 7. But perhaps if other parameters are changed, a mirror effect can be obtained with $n(A) = n(B)$.

Suppose subjects cannot discriminate A and B items or do not pay more attention to A than B items; can a mirror effect still be obtained if $n(A) = n(B)$? It turns out that it can if

Table 6
*Mirror Effect for Different Parameter Values Where,
in All Cases, $n(AO) = n(AN) = 60$, $n(BO) = n(BN) = 40$*

	Example			
	1	2	3	4
$p(A, old)$	.154	.154	.15	.16
$p(B, old)$	.136	.154	.15	.14
$p(B, new)$	.10	.12	.10	.12
$p(A, new)$	.10	.10	.10	.10
BO/BN	.69	.67	.75	.60
BO/AN	.78	.77	.78	.76
AO/BN	.75	.75	.77	.77
AO/AN	.81	.81	.80	.84
AO/BO	.63	.65	.55	.73
BN/AN	.66	.68	.57	.73

Note. A = low-frequency items; B = high-frequency items; O = old items; N = new items.

Table 7
*Mirror Effect for Different Parameter Values Where,
in All Cases, $n(AO) = n(AN) = n(BO) = n(BN) = 50$*

	Example			
	1	2	3	4
$p(A, old)$	.154	.154	.15	.16
$p(B, old)$	.136	.154	.15	.14
$p(B, new)$	.10	.12	.10	.12
$p(A, new)$	.10	.10	.10	.10
BO/BN	.71	.69	.77	.62
BO/AN	.82	.80	.77	.80
AO/BN	.80	.80	.77	.79
AO/AN	.79	.79	.77	.81
AO/BO	.64	.68	.50	.75
BN/AN	.67	.70	.50	.77

Note. A = low-frequency items; B = high-frequency items; O = old items; N = new items.

other parameters are changed. Assume that $p(BN) > p(AN)$ (new high-frequency words could well have more marked features than new low-frequency words). If $n(A) = n(B) = 50$, $N(B) = 1,000$, $N(A) = 800$, $p(AN) = .08$, and $p(BN) = .22$, then the forced-choice probabilities are .68, .78, .79, .82, .65, and .68 for BO/BN, BO/AN, AO/BN, AO/AN, AO/BO, and BN/AN, respectively. The feature distributions and the transformed binomial distributions rescaled to the LLRs are shown in Figures 7 and 8.[6]

Thus, the assumption that subjects pay more attention to A items than to B items does not seem to be necessary to generate a mirror effect. Obviously, there are trade-offs among the six parameter values, so probably other combinations would work, too. Glanzer et al. (1993) were clearly aware of this; they provided a detailed discussion of some possible modifications of the model that would give similar results (pp. 563–565). However, they did not discuss the case where $N(B) > N(A)$ or $n(A) = n(B)$.[7]

In an early article, Glanzer and Bowles (1976) gave a slightly different interpretation of the mirror effect for low- and high-frequency words. Among other things, they assumed that high-frequency words have more different meanings (S) than have low-frequency words, but subjects sample a fixed number (s) of the total possible meanings, and s is the same for low- and high-frequency words. This seems to be equivalent to saying that subjects attend to an equal number of features of low- and high-frequency words, but this

---

[5] In the fourth example, the null choices are too high, but in some cases subjects may know both items are old (or new) and guess.

[6] The means for the rescaled distributions are $-.802$, $-.205$, $.213$, and $.940$, and the variances are 1.352, 0.395, 0.441, and 2.179 for AN, BN, BO, and AO, respectively.

[7] Glanzer et al. (1993) did discuss at length the case where $p(i, old)$ is the average of $p(A, old)$ and $p(B, old)$, which they called $p(. old)$. This could come about in various ways. However, this is the deviation of the subject's model from the actual situation; if I understand it correctly, the actual situation is such that $n(A) > n(B)$, so subjects are still paying more attention to A items than to B items.
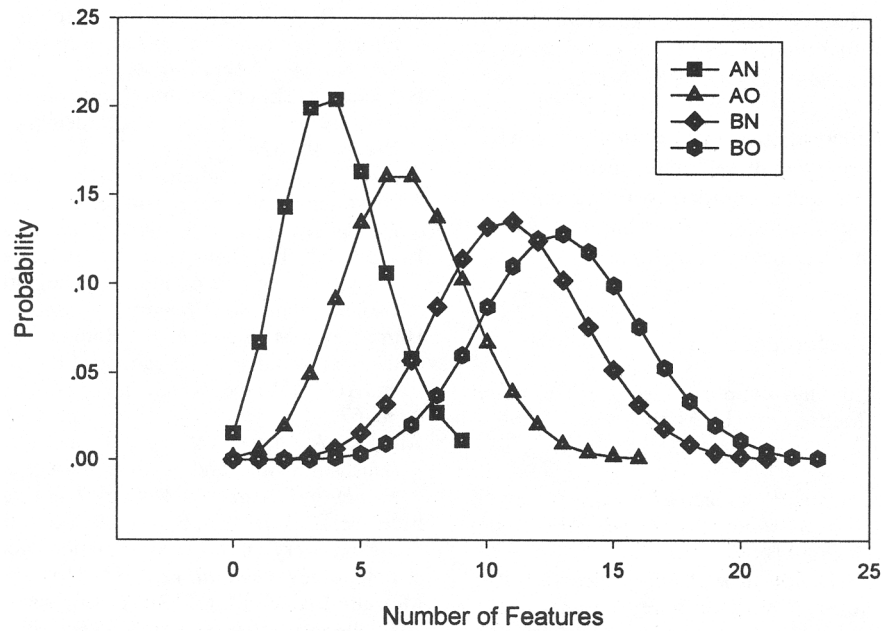
Figure 7. Binomial strength distributions with equal attention for AN, BN, BO, and AO. AN = low-frequency new items; BN = high-frequency new items; BO = high-frequency old items; AO = low-frequency old items.

assumption was changed in ALT, so now it is assumed that $n(A) > n(B)$.

Where does this leave us? If one wants to adopt the detailed assumptions of ALT, then clearly the theory works as claimed. However, other distributions (the Poisson and the hypergeometric) can produce a mirror effect and, even within the binomial framework, a variety of combinations of parameter values will generate a mirror effect.

Obviously, this does not mean that ALT is wrong; all current theories probably have variants that will also explain
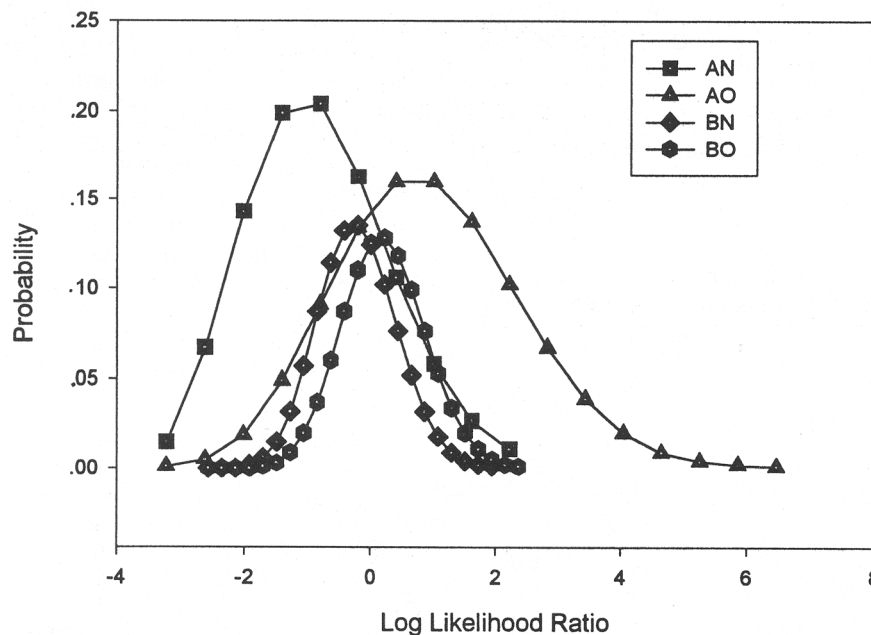

Figure 8. Transformed binomial distributions with equal attention rescaled to the log likelihood ratios. AN = low-frequency new items; BN = high-frequency new items; BO = high-frequency old items; AO = low-frequency old items.

at least some of the effects covered by the original theory. However, it does mean that the extensive empirical work of Glanzer and his colleagues (Glanzer & Adams, 1985; Glanzer et al., 1991; Glanzer et al., 1993; Kim & Glanzer, 1993) has not yet provided unequivocal evidence for ALT. There are competing interpretations, and further experiments are needed to test these competing interpretations. This discussion may focus the issues somewhat so a clearer understanding of the mirror effect can be attained.

References

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

Estes, W. K. (1959). The statistical approach to learning theory. In S. Koch (Ed.), *Psychology: A study of a science* (Vol. 2, pp. 145–154). New York: McGraw-Hill.

Feller, W. (1968). *An introduction to probability theory and its applications.* (3rd ed., Vol. 1). New York: Wiley.

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., & 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., Adams, J. K., Iverson, G. J., & Kim, K. (1993). The regularities of recognition memory. *Psychological Review, 100,* 546–567.

Glanzer, M., Kim, K., & Adams, J. K. (in press). Response distribution and the mirror effect. *Journal of Experimental Psychology: Learning, Memory, and Cognition.*

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.

Greene, R. L. (1996). Mirror effect in order and associative information: Role of response strategies. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 22,* 687–695.

Hintzman, D. L. (1988). Judgments of frequency and recognition memory in a multiple-trace memory model. *Psychological Review, 95,* 528–551.

Hintzman, D. L., Caulton, D. A., & Curran, T. (1994). Retrieval constraints and the mirror effect. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 20,* 275–289.

Hockley, W. E., Hemsworth, D. H., & Consoli, A. (in press). Shades of the mirror effect: Recognition of faces with and without sunglasses. *Memory & Cognition.*

Humphreys, M. S., Pike, R., Bain, J. D., & Tehan, G. (1989). Global matching: A comparison of the SAM, Minerva II, Matrix, and TODAM models. *Journal of Mathematical Psychology, 33,* 36–67.

Kim, K., & Glanzer, M. (1993). Speed versus accuracy instructions, study time, and the mirror effect. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 19,* 638–652.

McElree, B., & Dosher, B. A. (1993). Serial retrieval processes and the recovery of order information. *Journal of Experimental Psychology: General, 122,* 291–315.

Metcalfe-Eich, J. M. (1982). A composite holographic associative recall model. *Psychological Review, 89,* 627–661.

Murdock, B. B. (1970). Short-term memory for associations. In D. A. Norman (Ed.), *Models of human memory* (pp. 285–304). New York: Academic Press.

Murdock, B. B. (1982). A theory for the storage and retrieval of item and associative information. *Psychological Review, 89,* 609–626.

Murdock, B. B., & Anderson, R. E. (1975). Encoding, storage, and retrieval of item information. In R. L. Solso (Ed.), *Information processing and cognition: The Loyola symposium* (pp. 145–194). Hillsdale, NJ: Erlbaum.