Journal of Experimental Psychology:
Learning, Memory, and Cognition
1994, Vol. 20, No. 1, 201–205

Copyright © 1994 by the American Psychological Association, Inc.
0278-7393/94/$3.00

OBSERVATIONS

On Explaining the Mirror Effect

Douglas L. Hintzman

Two points are made in relation to the recent article by K. Kim and M. Glanzer (1993). First, the attention-likelihood model is more complex than these authors and others suggest. In particular, 2 kinds of quantities—(a) parameters representing the true state of the subject's memory and (b) the subject's estimates of those parameters—have been referred to using the same symbols. This obscures the essential role of metamemory in the model's predictions. Second, log-likelihood rescaling is not needed to explain the mirror effect. An alternative rescaling scheme is described, which can be added to a variety of memory models. This new rescaling method estimates a test item's learnability by learning it. Simulations show that the method is consistent with Kim and Glanzer's experimental results.

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

The mirror effect in recognition memory is the tendency for variables to affect hit and false-alarm rates in opposite directions. Thus, if correct "old" responses are more common in Condition A than in Condition B, incorrect "old" responses will be less common in Condition A than in Condition B. Interest in this phenomenon once focused almost exclusively on differences between high- and low-frequency words; however, Glanzer and Adams (1985, 1990) have argued convincingly that many variables can produce a mirror effect. They have also pointed out that, in the context of a signal-detection analysis of recognition decisions, the mirror effect implies that the underlying distributions are ordered as follows: new A < new B < old B < old A.

The attention-likelihood model, as developed and applied to data by Glanzer and his colleagues (Glanzer & Adams, 1990; Glanzer, Adams, & Iverson, 1991; Kim & Glanzer, 1993), offers an approach to understanding why distributions might be ordered in this way. Less obviously, it suggests an answer to the question of how recognition criteria are set. Both of these problems have proved difficult for memory theorists to solve.

If important lessons are going to be extracted from the attention-likelihood model, however, as much clarity as possible is needed about what assumptions it makes and how its predictions are derived. In this comment, I first clarify some aspects of the model that are obscure in Kim and Glanzer's (1993) article and in previous articles (Glanzer & Adams, 1990; Glanzer et al., 1991). I then describe an alternative way to explain Kim and Glanzer's results, without assuming that likelihoods are computed or known.

The Attention-Likelihood Model

The attention-likelihood model assumes that each item consists of N features and that prior to learning, a certain proportion, p_{new}, of those features are "marked."[1] During study, the proportion of marked features is increased according to the expression,

$$p_{\text{old}}(i,\text{study}) = p_{\text{new}} + q_{\text{new}} \cdot \left[\frac{n(i,\text{study})}{N}\right]. \tag{1}$$

The greater $n(i,\text{study})$ is—that is, the more attention is paid to items of type i—the more of the item's features become marked.

At retrieval, $n(i,\text{test})$ features of the test item are sampled at random. The number of these features that are marked is x, which constitutes a strength measure. The distribution of x is binomial. If the test item is new, the distribution of x is

$$P(x\,|\,\text{new}) = \binom{n(i,\text{test})}{x} \cdot p^{x}_{\text{new}} \cdot q^{n(i,\text{test})-x}_{\text{new}}. \tag{2}$$

If the test item is old, the distribution is

$$P(x\,|\,\text{old}) = \binom{n(i,\text{test})}{x} \cdot p_{\text{old}}(i,\text{test})^{x} \cdot q_{\text{old}}(i,\text{test})^{n(i,\text{test})-x}. \tag{3}$$

Here, $p_{\text{old}}(i,\text{test})$ indicates that $p_{\text{old}}(i,\text{study})$ has been decremented as a result of forgetting.[2] When there has been more forgetting in one condition than in another, $p_{\text{old}}(i,\text{test})$ for the two conditions will be different, even if $p_{\text{old}}(i,\text{study})$ was the same. For other experiments, such as Kim and Glanzer's (1993) Experiment 2, $p_{\text{old}}(i,\text{test})$ differs between conditions because $p_{\text{old}}(i,\text{study})$ was different, even though the proportion forgotten has been the same. Note that Equations 2 and 3 are

This article is based on research supported by National Science Foundation Grant BNS-90-08909.

Correspondence concerning this article should be addressed to Douglas L. Hintzman, Department of Psychology, University of Oregon, Eugene, Oregon 97403. Electronic mail may be sent to hintzman@oregon.uoregon.edu.

[1] This treatment of the attention-likelihood model adopts slightly different notations from the one previously used (Glanzer & Adams, 1990; Glanzer, Adams, & Iverson, 1991; Kim & Glanzer, 1993). The primary differences are that I consistently differentiate between study and test parameters in parentheses, move "old" and "new" from parentheses to subscripts, and consistently use q in place of $1 - p$. As shall be described, I also distinguish subjects' estimates of parameters from the parameters themselves.

[2] Glanzer, Adams, and Iverson (1991, Equation 4) referred to this as $p(i,\text{old},t)$, where t was the amount of time since study.

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

derived strictly from assumptions about learning and retention and thus represent the true state of the subject's memory at the time of the test.

When an x value is determined for a test item, the subject must decide whether that count came from a new item or from an old item, that is, from Equation 2 or Equation 3. Attention-likelihood theory proposes that the subject decides this by determining the likelihood ratio for x: $P(x|\text{old})/P(x|\text{new})$. This ratio has the value 1 at the point where the new and old distributions cross, and $\ln(1) = 0$. Hence, if x is rescaled as $\ln[P(x|\text{old})/P(x|\text{new})]$, all pairs of corresponding old and new distributions will be aligned at their crossing points.[3] This rescaling is sufficient to predict the mirror effect, and the value 0 on this scale is a natural place to set a neutral recognition criterion because hits and correct rejections are equally probable at this point.

How does the subject compute the likelihood ratio for x? Equations 2 and 3 represent the true state of the subject's memory. To compute a likelihood ratio, the subject must have a way of estimating the true state. Kim and Glanzer (1993) acknowledged this need for estimates but nevertheless used the symbols for the true states in their rescaling equation (their Equation 2). To clarify that rescaling must be based on estimates of the relevant parameters, rather than on the parameters themselves, I rewrite the rescaling equation as follows:

$$\ln L(x) = n(i,\text{test}) \cdot \ln\left[\frac{\hat{q}_{\text{old}}(i,\text{test})}{\hat{q}_{\text{new}}}\right] + x \cdot \ln\left[\frac{\hat{p}_{\text{old}}(i,\text{test}) \cdot \hat{q}_{\text{new}}}{\hat{p}_{\text{new}} \cdot \hat{q}_{\text{old}}(i,\text{test})}\right]. \quad (4)$$

L(x) is the likelihood ratio for x, and the "^" indicates the subject's estimate of the appropriate parameter (p or q). Equation 4 describes a straight line with a negative intercept (the product of $n(i,\text{test})$ and a negative-valued logarithm) and a positive slope (the logarithm that multiplies x).

It is important to be clear that parameters of the true state of the subject's memory (p_{new} or p_{old}), on the one hand, and estimates of those parameters ($\hat{p}_{\text{new}}$ and $\hat{p}_{\text{old}}$), on the other, are different concepts. By using the same symbols for both, Kim and Glanzer (1993) made the attention-likelihood model seem simpler than it is. Their notation implies that subjects have perfect knowledge of their memory states, and it explicitly sanctions substitution of the true values, from Equations 2 and 3, for the estimates of those values in Equation 4. Failing to distinguish between the parameters and their estimates also obscures several interesting questions concerning how subjects compute the estimates, the accuracy of the estimates, whether subjects might be misled about their memory states, and how such misconceptions might affect recognition performance. Always using the true states in Equation 4 also constrains the model to predict a mirror effect for any manipulation that affects recognition performance, even though mirror effects do not always occur (e.g., Hoshino, 1991; Murnane & Phelps, 1993; Wixted, 1992).

These seem to be central questions. Assuming that subjects have exactly the right rescaling equation to produce a mirror effect (Equation 4) and that they know exactly the right parameter values to enter into the equation comes perilously close to circularity. Glanzer and his colleagues (Glanzer & Adams, 1990; Glanzer et al., 1991; Kim & Glanzer, 1993) have shown that log-likelihood rescaling would be sufficient to produce a mirror effect but have offered no explanation of how subjects accomplish the rescaling. This is a major, unacknowledged gap in their account of the mirror effect.

Kim and Glanzer (1993) presented two tests of the attention-likelihood model. In Experiment 1, they tested some subjects under accuracy instructions and other subjects under speed instructions. This manipulation directly affects $n(i,\text{test})$ and indirectly affects x in Equations 2–4. Because fewer features are sampled under speed instructions than under accuracy instructions, all discriminations are impaired. In Kim and Glanzer's Experiment 2, the study condition was varied (long vs. short study time) and the test condition was held constant. Study time affects $p_{\text{old}}(i,\text{study})$, which scales down to a difference in $p_{\text{old}}(i,\text{test})$ in Equation 3. Thus, longer study times tend to yield higher values of x for old items. However, the most interesting outcome of the experiment involved new items: In a forced-choice task, the tendency to choose a new high-frequency (HF) word over a new low-frequency (LF) word was smaller when the study time had been short than when the study time had been long. This is the outcome that Kim and Glanzer called "concentering."

The concentering of new-item distributions cannot be explained by Equation 2 because study time affects the true memory states of only old items and not those of new ones. The result can only arise from Equation 4. In that equation, $n(i,\text{test})$ is assumed to reflect only test conditions, and $p_{\text{old}}(i,\text{test})$ does not appear. Kim and Glanzer (1993) circumvented this problem by failing to distinguish the true-state parameter from its estimate, that is, by substituting $p_{\text{old}}(i,\text{test})$ for $\hat{p}_{\text{old}}(i,\text{test})$ in Equation 4. Glanzer et al. (1991) similarly found concentering of new items as a result of forgetting (the forced-choice preference for new HF words over new LF words declined when the retention interval was long), and that result was similarly modeled by using $p_{\text{old}}(i,\text{test})$ in Equation 4.

It is reasonable to believe, as Kim and Glanzer (1993) argued, that subjects can remember whether the presentation rate for the preceding list was long or short and that they know that slow rates lead to better learning than fast rates do. It is likewise reasonable to suppose subjects are aware that memory declines over time. Such metacognitive judgments may well play a role in recognition memory and could do so by figuring into estimates such as $\hat{p}_{\text{old}}(i,\text{test})$,in Equation 4. It is nevertheless essential that theorists make it clear when they are invoking such judgments, by symbolically distinguishing subjects' estimates of quantities from the quantities themselves. Making this distinction also reveals the model's implicit assumption of high-level cognitive operations that are not explained. Such explanatory gaps are present in all memory models (including

[3] Kim and Glanzer's (1993) Equation 3 combined both of the present Equations 2 and 3, after x was transformed to the log-likelihood scale.

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

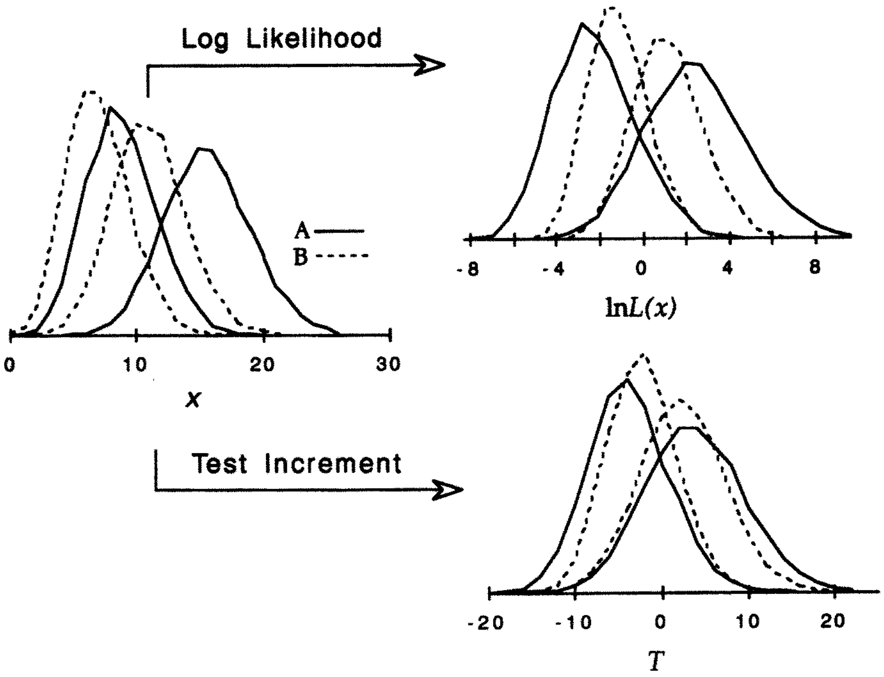

Figure 1. Effects of two rescaling schemes applied to the basic learning and retrieval assumptions of Glanzer and Adams's (1990) model. On the left are the raw distributions of the number of marked features retrieved, x (the strength distributions). On the top right are the results of log-likelihood [ln$L(x)$] rescaling according to Equation 4. On the bottom right are the results of test-trial increment (T) rescaling, according to Equation 5.

the one that follows), but if models are to promote understanding, their gaps need to be identified.

Rescaling by Test-Trial Increments

The second point I wish to make in this commentary is that strengths can be rescaled to produce a mirror effect without resorting to likelihood ratios. After x is transformed by Equation 4, corresponding pairs of new and old distributions straddle the new origin so that they are lined up at their crossing points. The expressions for intercept and slope in Equation 4 are specifically tailored for binomial strength distributions of the type generated by the learning and retrieval assumptions of Glanzer and Adams's (1990) model and would be useless for strength distributions generated by some other process. However, more generally applicable transformations might be accomplished in other ways. For example, a transformation could be based on estimates of the means of the new and old distributions and estimates of one or both standard deviations, in effect turning retrieved strengths into z scores (cf. Gillund & Shiffrin, 1984).

A simpler scheme approximates the midpoint between the new- and old-distribution means and places the origin there. This can be accomplished if one has an estimate of the mean of the new-item distribution (the baseline) and an estimate of the increment that would have been produced by a study trial on the test item. Such estimates might be obtained in a variety of ways. Perhaps the most straightforward scheme for estimating an item's learning increment is to learn the item at the time of test: First assess the test item's retrieval strength, then learn the item, and then assess its retrieval strength again. If each item has a characteristic learning rate (e.g., generally higher for LF than for HF items), the test-trial increment should be a good estimate of the study-trial increment. In this way, the item presented for test can, in effect, rescale itself. This scheme, which can be called test-trial increment rescaling, avoids the necessity of keeping track of learning rates for many different categories of items (e.g., Glanzer & Adams, 1990).

In principle, test-trial increment rescaling can be added to any memory model that allows learning to occur on a test trial. For present purposes, I illustrate how it works by applying it to the basic learning and retrieval assumptions of the Glanzer and Adams (1990) model, which is a version of stimulus-sampling theory (Estes, 1950). Figure 1 shows the results of simulations for two categories of items, A and B, that have different learning rates (e.g., LF and HF words). The parameters of the simulation were as follows: $p_{\text{new}} = .1$, $N = 1000$, $n(A,\text{study}) = n(A,\text{test}) = 90$, and $n(B,\text{study}) = n(B,\text{test}) = 70$. The left panel shows the distributions of x for new and old items of Types A and B. These are the distributions of raw strengths described by Equations 2 and 3. The top right panel shows the result of log-likelihood rescaling using Equation 4 and assuming $\hat{p}_{\text{new}} = p_{\text{new}}$ and $\hat{p}_{\text{old}}(i,\text{test}) = p_{\text{old}}(i,\text{test})$.

The bottom right panel shows the effect of rescaling the same x values using their test-trial increments. The rescaling was done as follows: On a test, $n(i,\text{test})$ features are sampled, and the number marked is counted (call this x_1). All the

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

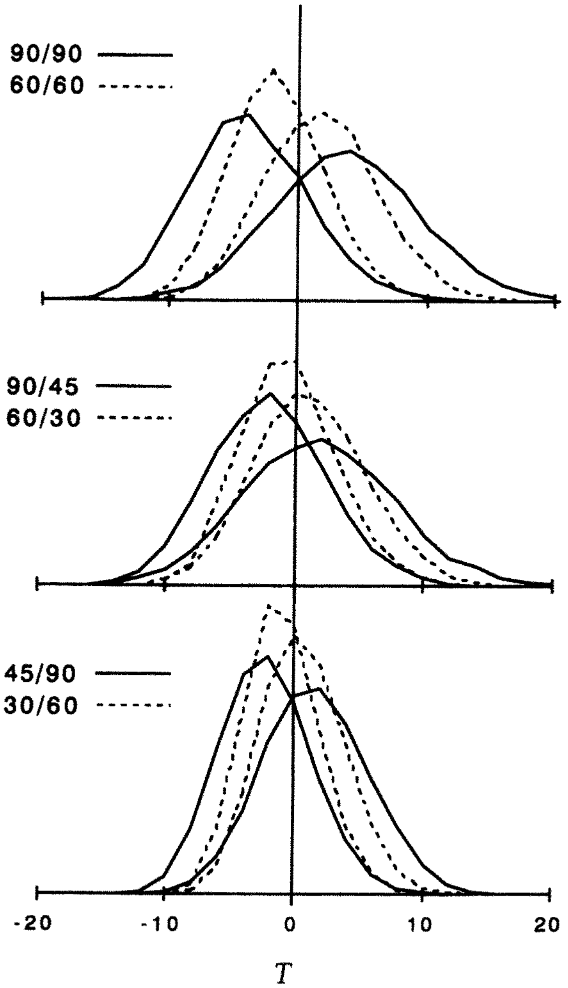

Figure 2. Rescaled distributions of x using the test-trial increment (T) scheme for three conditions tested by Kim and Glanzer (1993). The top panel shows the condition $n(A,\text{study}) = n(A,\text{test}) = 90$, $n(B,\text{study}) = n(B,\text{study}) = 60$, and $k = .5$. The middle panel shows the condition $n(A,\text{study}) = 90$, $n(A,\text{test}) = 45$, $n(B,\text{study}) = 60$, $n(B,\text{study}) = 30$, and $k = 1$. The bottom panel shows the condition $n(A,\text{study}) = 45$, $n(A,\text{test}) = 90$, $n(B,\text{study}) = 30$, $n(B,\text{study}) = 60$, and $k = .25$.

sampled features are now marked (the test-trial learning phase), and a second random sample of size $n(i,\text{test})$ is taken. The number of features in this second sample that are marked is counted, yielding x_2; thus, the test-trial increment is $\Delta x = x_2 - x_1$. The initial count, x_1, can now be rescaled to the value T as follows:

$$T = x_1 - [k \cdot \Delta x + \hat{p}_{\text{new}} \cdot n(i,\text{test})]. \qquad (5)$$

Like Equation 4, this equation describes a linear transformation. The expression in brackets is the (negative) intercept. The second term in the brackets $\hat{p}_{\text{new}} \cdot n(i,\text{test})$ is the estimated mean of the new-item distribution. or baseline. The first term in the brackets is the test-trial increment (Δx) multiplied by a parameter, k, that locates the origin relative to the new and old distributions. If the degree of learning at test is the same as that at study (i.e., if $n(i,\text{test}) = n(i,\text{study})$), then $k = 0$ places the origin at the mean of the new-item distribution, and $k = 1$ places it at the mean of the old-item distribution. One can put the origin halfway between the two means by setting k at .5.[4] The data in the bottom right panel of Figure 1 were generated with $k = .5$ and $\hat{p}_{\text{new}} = p_{\text{new}}$.

To determine whether this model is consistent with the findings of Kim and Glanzer (1993), I performed simulations of three conditions. Each simulation run included learning and testing on 10,000 items, and the parameter N was set at 1,000. For one run, simulating long study times and accuracy instructions at retrieval, $n(\text{A,study}) = n(\text{A,test}) = 90$, $n(\text{B,study}) = n(\text{B,test}) = 60$, and $k = .5$. The transformed distributions produced by this simulation are shown in the top panel of Figure 2. The second simulation was for long study times and speed instructions at retrieval. It used $n(\text{A,study}) = 90$, $n(\text{A,test}) = 45$, $n(\text{B,study}) = 60$, $n(\text{B,test}) = 30$, and $k = 1$, and the results are shown in the middle panel of Figure 2. The third simulation was for short study times and accuracy instructions at retrieval, and used $n(\text{A,study}) = 45$, $n(\text{A,test}) = 90$, $n(\text{B,study}) = 30$, $n(\text{B,test}) = 60$, and $k = .25$. These results are shown in the bottom panel of Figure 2.

The parameter k can be seen as embodying certain aspects of metacognition: k is high if the subject thinks test-trial learning is poorer than study-trial learning was, and it is low if the subject believes that test-trial learning is better than study-trial learning or (as in Glanzer et al., 1991) that there has been significant forgetting prior to test. Thus in its relation to the subject's knowledge of the experimental situation, k serves a function in this model similar to that of $\hat{p}_{\text{old}}(i,\text{test})$ in Equation 4. It differs from $\hat{p}_{\text{old}}(i,\text{test})$, however, in requiring only a relativistic judgment (the new test increment vs. the present state of the original study increment) and not an absolute value. (The absolute estimate of degree of learning is given by Δx.) The concept of $\hat{p}_{\text{new}}$ is the same in both rescaling schemes. Thus the new model has the same number of parameters as the old one ($\hat{p}_{\text{old}}(i,\text{test})$ is dropped and k is added). An obvious way to test this new model is to try to fool subjects into using an inappropriate value of k.

The simulated data displayed not only the mirror effect, but also the changes in overlap of new A and new B distributions as a function of study time called concentering (Glanzer et al., 1991; Kim & Glanzer, 1993). Forced-choice proportions computed from the distributions shown in Figure 2 show a preference of new B > new A of .644 in the top panel and .606 in the bottom panel (based on 10,000 observations, the standard error of the proportion was .005). The only difference between the two simulations that can account for this difference in outcomes is the rescaling parameter, k.

Conclusion

I have argued that attention-likelihood theory contains hidden parameters representing subjects' metamemorial judgments and that the notation used to date has understated the

[4] This description assumes that the baseline estimate is correct, that is, $\hat{p}_{\text{new}} = p_{\text{new}}$. The calibration of k with respect to the new and old distributions will be thrown off by systematic errors in $\hat{p}_{\text{new}}$.

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

model's complexity by obscuring this fact. I have also demonstrated an alternative way to rescale strengths without resorting to log-likelihood ratios. In this scheme, learning on the test trial is combined with a metamemorial judgment about the relationship between the situations at study and test, to arrive at an adjustment of the strength-scale origin. This approach describes a mechanism—not just an equation—for rescaling and clearly distinguishes states of the memory system from judgments about those states.

An advantage of this new approach to explaining the mirror effect is that it can be added to virtually any memory model that allows learning to occur on test trials. To facilitate comparison of the two rescaling schemes, I have shown how the new one can be added to the feature-sampling model used by Glanzer and his colleagues. Certain complexities in the rescaling scheme as presented here arise from its interface with this particular model, for example, the requirement that two independent samples of features be taken in quick succession on a test trial and the necessity of computing and subtracting the baseline estimate, in Equation 5.

Rescaling by test-trial increments appears less complex when added to a model that has a baseline of zero, such as Minerva 2 (Hintzman, 1988) or TODAM (Murdock, 1982). For example, a version of Equation 5 appropriate for the Minerva 2 model is: $T = I_1 - k \cdot \Delta I$, where I_1 is the initial echo intensity or strength produced by the test item and $\Delta I = (I_2 - I_1)$ is the increment in echo intensity due to learning on the test trial. A point of caution is that it is not necessarily safe to assume that rescaling by test-trial increments will behave identically when added to different memory models because the learning, retrieval, and rescaling components of models interact.

Readers attracted by the seeming elegance of attention-likelihood theory may be chagrined by the present revelation of its hidden assumptions and put off by the complexity of the alternative I have offered. The elaborateness of both schemes appears to derive from their starting with the standard view that study increases strength or familiarity and then adding a way to reorder strengths so as to produce a mirror effect. A more appealing account of the mirror effect may require a different approach. The challenge is to come up with that account.

References

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

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.

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

Hoshino, Y. (1991). A bias in favor of the positive response to high-frequency words in recognition memory. *Memory & Cognition, 19,* 607–616.

Kim, K., & Glanzer, M. (1993). Speed versus accuracy instructions, study time, and the mirror effect. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 19,* 638–654.

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

Murnane, K., & Phelps, M. P. (1993). A global activation approach to the effect of changes in environmental context on recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 19,* 882–894.

Wixted, J. T. (1992). Subjective memorability and the mirror effect. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 18,* 681–690.

Received September 30, 1992
Revision received February 8, 1993
Accepted February 15, 1993 ■