Psychological Review
2004, Vol. 111, No. 3, 588–616

0033-295X/04/$12.00 DOI: 10.1037/0033-295X.111.3.588

Sum–Difference Theory of Remembering and Knowing: A Two-Dimensional Signal-Detection Model

Caren M. Rotello, Neil A. Macmillan, and John A. Reeder
University of Massachusetts

In the remember–know paradigm for studying recognition memory, participants distinguish items whose presentations are episodically remembered from those that are merely familiar. A one-dimensional model postulates that remember responses are just high-confidence old judgments, but a meta-analysis of 373 experiments shows that the receiver operating characteristic (ROC) curves predicted by this model have the wrong slope. According to the new *s*um–difference *t*heory of *re*membering *a*nd *k*nowing (STREAK), old items differ from new ones in both global and specific memory strength: The old–new judgment is based on a weighted *sum* of these dimensions, and the remember–know judgment is based on a weighted *difference*. STREAK accounts for the form of several novel kinds of ROC curves and for existing remember–know and item-recognition data.

Over the last 30 years, memory researchers have debated whether recognition judgments depend only on the global strength of memories or on global strength as well as a recall-like process. Single-process accounts, like the global matching models, assume that such decisions are based exclusively on an assessment of the *familiarity* of the test item (e.g., Gillund & Shiffrin, 1984; Hintzman, 1988). These memory models account for a wide variety of data (for reviews, see Clark & Gronlund, 1996; Raaijmakers & Shiffrin, 1992). A newer body of work, however, has persuaded many researchers to return to the more traditional view (e.g., Atkinson & Juola, 1974) that recognition judgments can also involve the *recollection* of the occurrence of the item, or aspects thereof (see Mandler, 1991, for a brief review). In this *two-process* view, the familiarity-based assessment of the test probe is supplemented by a recollective process that provides more specific information.

Empirical support for the role of recollection in recognition has come in a variety of forms. For example, familiarity seems to operate more quickly than recollection. Participants tested on foils that are similar to but different from study items incorrectly say "old" if forced to make rapid recognition decisions, apparently basing their judgments on the global familiarity of the probes. When allowed more time to make their judgments, however, they are increasingly likely to call those foils "new," consistent with the idea that recollected details become available with time (e.g., Dosher, 1984; Rotello & Heit, 2000). Evoked response potential (ERP) experiments have also provided support for the two-process view of recognition: Familiarity-based and recollective recognition responses are associated with temporally distinct patterns of activation in spatially distinct recording sites (see Allan, Wilding, & Rugg, 1998, for a review).

One way to distinguish the effects of familiarity and recollection in standard, unspeeded recognition is to compare performance across memory tasks in which these two processes act either in concert or in opposition. In the *process dissociation procedure* (Jacoby, 1991) and one recent extension, *conjoint recognition theory* (Brainerd, Reyna, & Mojardin, 1999), experimental participants identify previously studied items according to decision rules that depend either on familiarity and recollection together or on recollection alone. Manipulations of study and test conditions produce changes in the inferred contributions of familiarity and recollection, under some fairly strong (and controversial) assumptions (e.g., T. Curran & Hintzman, 1995; Ratcliff, Van Zandt, & McKoon, 1995).

Another method that has been used to assess the contributions of recollection and familiarity to recognition judgments is the *remember–know* (R-K) paradigm. First proposed by Tulving (1985), this paradigm is appealing in its simplicity: A single memory task and a single class of studied items are sufficient. In this article, we critically evaluate a signal-detection theory (SDT) model that provides the only current quantitative analysis of this paradigm (Donaldson, 1996; Hirshman & Master, 1997). A meta-analysis of the current literature shows that this one-dimensional model cannot account for the essential results it is intended to explain. We then offer a new two-dimensional model that, although still grounded in SDT, is more consistent with Tulving's original proposal. Our model accounts for the current database and also leads to novel experimental tests, which we report below, that support it.

The Remember–Know Paradigm

Participants in R-K experiments are asked whether they recognize previously studied words or pictures presented in a test list

Caren M. Rotello, Neil A. Macmillan, and John A. Reeder, Department of Psychology, University of Massachusetts.

John A. Reeder is now at the Department of Psychology, Simmons College.

Portions of the model and data were presented at the 26th Annual Interdisciplinary Conference, Jackson Hole, Wyoming, January 2001, and at the 42nd Annual Meeting of the Psychonomic Society, Orlando, Florida, November 2001. This research was supported by National Institutes of Health Grant R01 MH60274 to Caren M. Rotello and Neil A. Macmillan.

We thank Barb Chalfonte, Jerry Myers, and Gordon Van Tassel for their assistance during the development of this project; Andy Yonelinas for sharing his data; and Otilia Catro and G. Elizabeth Johnson for their assistance with data collection. Bill Banks, Barbara Dosher, Evan Heit, Ken Norman, Jeroen Raaijmakers, Michael Verde, and Andy Yonelinas provided helpful comments on an earlier version of this article.

Correspondence concerning this article should be addressed to Caren M. Rotello, Department of Psychology, University of Massachusetts, Box 37710, Amherst, MA 01003-7710. E-mail: caren@psych.umass.edu

that also includes unstudied words or pictures. For each item they identify as studied, they are also asked to evaluate the phenomenal basis of that decision: Do they remember something specific about the experience of studying the item, or do they feel as if they know they studied it, despite its failure to evoke conscious recollection of any details about the experience? A number of systematic differences in remember and know response rates have suggested that the judgments reflect different kinds of information processing (for a recent review, see Gardiner & Richardson-Klavehn, 2000). For example, uncommon words are remembered more often than common words, but they are known in equal proportions (e.g., Gardiner & Java, 1990). Such dissociations have led some researchers to interpret remember and know decisions as *process-pure* responses that reflect recollection and familiarity, respectively (e.g., Conway & Dewhurst, 1995; Gardiner & Java, 1991). Others have postulated implicit and explicit memory systems that are associated exclusively with the two kinds of retrieval experiences (e.g., Rajaram, 1993; Tulving, 1985).

Under this process-pure interpretation, the R-K paradigm directly estimates the contributions of recollection and familiarity to recognition and avoids the need for multiple tasks like those used in process dissociation and conjoint recognition theory. However, the paradigm does require participants to make multiple decisions, and the data are complex. The old–new choice is sometimes modified by confidence ratings, and both the old–new and R-K decisions can be influenced by response bias. Although the R-K distinction itself is subjective, both remembering and knowing can occur with correct and incorrect claims of recognition. A full account of R-K data must offer a theoretical description of the presumed underlying processes and the method of their combination in making these memory judgments. SDT is a natural framework within which to address issues of this sort.

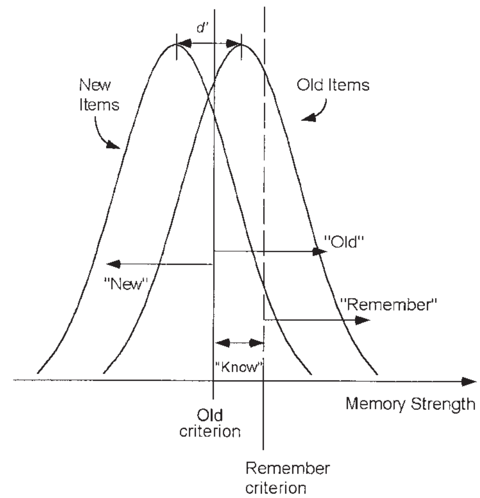

Figure 1. The one-dimensional signal-detection model of remember–know judgments. Old and New items differ in average strength, and two criteria are used to determine responses. Observations above the upper criterion lead to remember responses, those between the two criteria lead to know responses, and those below the lower criterion lead to new responses.

The One-Dimensional Model of Remember–Know Judgments

Donaldson (1996) and Hirshman and Master (1997) independently proposed a one-dimensional, familiarity-based SDT model to account for R-K data. The model is displayed graphically in Figure 1. All recognition decisions are based solely on the familiarity, or global memory strength, of the test item. On average, test items that are presented for study have greater memory strength than those that are novel to the experimental setting, but there is some overlap in the strength of the two classes of stimuli. To make a recognition decision, the participant establishes a criterial level of memory strength, calling all items with strength greater than that criterion "old" and all weaker items "new." To make the R-K distinction, the participant establishes a second, higher criterion. Items are "remembered" when their familiarity exceeds this high criterion, and they are "known" when it falls between the two criteria. In this view, then, remembering that a test item was previously encountered is not qualitatively different from knowing that it was; only a degree of memory strength or confidence distinguishes the experiences.

One empirical test of the one-dimensional model arises from its fundamental premise that remember and old responses reflect only a difference in memory strength or confidence. The distance between the means of the two distributions in Figure 1 is theoretically a measure of memory accuracy that is independent of the decision criterion when the variances of the two distributions are equal: The same distance should be found regardless of whether a high (i.e., remember) or low (i.e., old) decision criterion is used to generate the hit and false-alarm rates (Macmillan & Creelman, 1991). According to the model, remember responses reveal the same memory accuracy as remember and know responses combined.

Using this logic, Donaldson (1996) evaluated the one-dimensional model by reanalyzing data from 80 published R-K conditions, and Gardiner, Ramponi, and Richardson-Klavehn (2002) did the same for a set of 86 remember–know–guess conditions. They each computed two estimates of sensitivity, the d' statistic of SDT and the alternative accuracy measure A', for remember and old responses. A consistent finding emerged in these analyses: A' was slightly but reliably higher for old than for remember responses, and d' was slightly lower for old than for remember responses. The modest discrepancies are roughly consistent with the equal-variance one-dimensional model in that, on the average, measures of performance do not differ for the two response criteria.[1] In fact, using a larger database of about 400 conditions, Dunn (2004) concluded that d' was essentially identical at the remember and old criteria.[2]

Item-recognition receiver operating characteristics (ROCs). The one-dimensional model in Figure 1 is a natural extension of a model that is well-supported by a different experiment, the confidence-rating design, in which participants rate their confidence in the old–new judgment. The different levels of confidence

[1] Macmillan, Rotello, and Verde (2004) have shown that the apparent discrepancy in the conclusions based on A' and d' is a natural consequence of the properties of A' and d' and argued that d' is the preferred measure. The small decrease in d' for the more lenient response criterion is consistent with the Old distribution having a greater variance than the New distribution has. For elaboration of this point and for discussion of the implications of adding a guess response, see Macmillan et al. (2004).

[2] Dunn (2004) also argued that several other types of data previously taken as evidence against the one-dimensional model are actually consistent with it. However, he did not consider the two-point *z*ROC slope data that we describe below.

are interpreted as different criterion settings, so that a high-confidence old response corresponds to a high, conservative criterion and a high-confidence new response to a low, liberal one. Rating data are used to construct an ROC curve, in which the probability of correctly responding "old" (the hit rate) is plotted against the probability of incorrectly responding "old" (the false-alarm rate). Each level of confidence leads to a separate hit rate (the area under the Old distribution above the criterion) and false-alarm rate (the area under the New distribution above the same criterion). Two theoretical ROC curves are shown in Figure 2A; each sweeps out the predicted hit and false-alarm rates at all possible criterion placements, from the most stringent (represented in the lower left corner of Figure 2A, yielding few old responses to either Old or New items) to the most lenient (in the upper right, yielding an old response to essentially all stimuli). The upper curve contains higher hit rates than the lower one for the same false-alarm rates and thus indicates superior accuracy.

The ROC curve can be transformed into normal–normal (z) space by finding the z scores (one for the hit rate and one for the false-alarm rate) of each point in probability space. The resulting zROC, shown in Figure 2B, has some convenient properties. First, it is linear whenever the recognition decisions are based on normal distributions of familiarity. Second, the y-intercept of the zROC indicates d' when the distributions have equal variance. Finally, and most important for present purposes, the slope of the zROC is the ratio of the standard deviations of the New and Old distributions. In Figure 2B, the lower curve has a slope of 1.0, implying that the Old and New distributions have equal standard deviations, whereas the upper curve has a slope of 0.80, implying that the standard deviation of the Old distribution is greater than that of the New distribution.

A number of traditional old–new recognition experiments have assessed the slope of the zROC by using either the confidence rating design or changes in instructions or payoffs (e.g., Glanzer, Kim, Hilford, & Adams, 1999; Hirshman & Hostetter, 2000; Ratcliff, Sheu, & Gronlund, 1992). Although these experiments were designed to manipulate slope over as wide a range as possible, the obtained slopes have been quite consistent: The median slope has ranged from 0.74 to 0.83, and the overall mean is about 0.80. In other words, the Old item distribution has almost always been found to have a standard deviation that is larger than that of the New item distribution. A histogram constructed from 103 zROC slopes reported by Ratcliff et al. (1992), Glanzer et al. (1999; who included a summary of zROC slopes for 21 other experiments), and Hirshman and Hostetter (2000) is shown in Figure 3A.

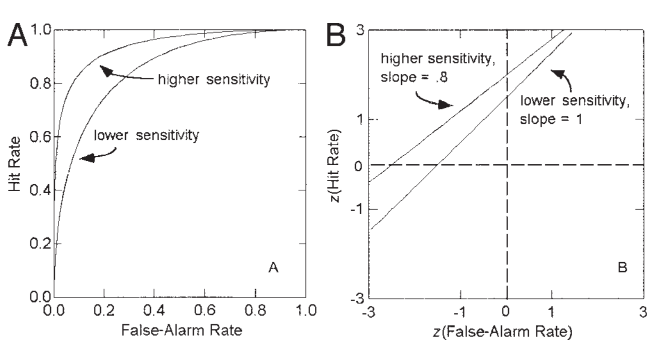

Figure 2. Sample receiver operating characteristic curves in probability space (A) and z space (B). The hit rate is the proportion of old responses to Old items, and the false-alarm rate is the proportion of old responses to New items.

Two-point ROCs. Because the one-dimensional R-K model assumes that remember and old responses are merely high- and low-confidence recognition decisions, it predicts that zROC slopes computed from the hit and false-alarm rates for those two confidence levels will be equal to slopes computed from confidence rating experiments. We can construct a *two-point ROC curve* from R-K data, the lower point representing the remember hit and false-alarm rates and the upper point indicating the old hit and false-alarm rates. It is important to realize that, although the curve is plotted on ROC coordinates, it cannot be given its usual interpretation as an iso-sensitivity contour (Luce, 1963) unless the one-dimensional model is correct. For that model to be consistent with the literature on old–new recognition, it must predict the slope of this curve (on z coordinates) to be about 0.80, just as though the criterion had been manipulated by instructions, payoffs, or confidence ratings. To test this prediction, we reanalyzed the 80 R-K experimental conditions reported by Donaldson (1996) and 293 others that have been published since 1996.[3] In each experimental condition, the slope of the two-point zROC was found as

$$z\text{ROC slope} = \frac{z(\text{HR}) - z(\text{remember HR})}{z(\text{FAR}) - z(\text{remember FAR})}, \tag{1}$$

where z(remember HR) and z(remember FAR) are the normalized hit and false-alarm rate resulting from the R-K criterion and z(HR) and z(FAR) are the normalized hit and false-alarm rate resulting from the old–new criterion. The resulting two-point zROC slopes are shown in summary form in Figure 3B and in Appendix A for each experimental condition.

The two-point R-K slope data in Figure 3B are strikingly different from the recognition zROC slopes in Figure 3A. First, the variance of two-point zROC slopes is substantially larger. Indeed, the upper tail of the distribution includes slopes larger than those observed in any published recognition study, even though old–new recognition experiments have been designed to produce changes in zROC slope. Second, the mean of the slopes is 1.01, and the median is 0.92; both are greater than is observed in the recognition data.

We considered possible artifactual explanations for the differences in variance and in central tendency between item-recognition and two-point ROC slopes. One potential confound is that ROC curves with fewer operating points produce slope estimates with greater standard errors (Macmillan, Rotello, & Miller, in press): Can this difference account for the variance discrepancy? To find out, we selected two points from each of the five-point recognition ROC curves based on group data reported by Ratcliff et al (1992): the highest confidence point (most comparable to the remember criterion in the one-dimensional model) and the middle point (most comparable to the old criterion). The resulting *two-*

[3] These 293 conditions included all of the published (and none of the unpublished) remember–know–guess studies summarized by Gardiner et al. (2002). Also, a number of experimental conditions yielded no false alarms. When the false-alarm rate equals 0, z(false-alarm rate) is undefined and the two-point slope cannot be calculated. All such cases were excluded from these analyses.

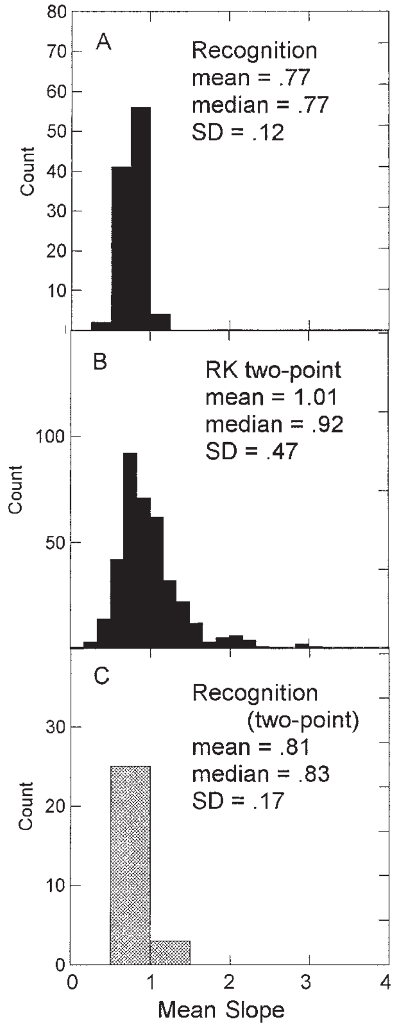

Figure 3. A: Distribution of normal–normal receiver operating characteristic (*z*ROC) slopes from the recognition literature ($N = 103$). B: Distribution of two-point *z*ROC slopes from the remember–know (RK) literature calculated using Equation 1 ($N = 373$). C: Distribution of two-point recognition *z*ROC slopes ($N = 28$; see the text for details). The one-dimensional model predicts that the mean of the recognition and RK slopes should be equal and that the variance of the two kinds of two-point slopes should be equal.

point recognition slopes, shown in Figure 3C, have a slightly greater standard deviation than those based on five points (0.17 vs. 0.12) but are still much less variable than the two-point R-K slopes ($SD = 0.47$). The latter variance difference is reliable: $t(73.9) = 4.615$, $p < .001$, Brown–Forsythe test. We conclude that the variance of the two-point R-K slopes is inflated by some other factor, such as a process that has been left out of the one-dimensional model.

The difference in the means (and medians) of the two-point R-K *z*ROCs and the recognition *z*ROC slopes could be a consequence of the skew present in the distribution of two-point slopes, but two calculations eliminate this possibility. First, trimming the two-point slopes that are more than two standard deviations from the mean, or that are in the most extreme 10% of each tail, does not destroy the effect. The trimmed means are 0.93 and 0.94, respectively, and the trimmed medians are 0.89 and 0.92; all are larger than the old–new mean and median (0.77 for both). Second, in a variety of simulations with the one-dimensional model, the two-point slopes were found to be unbiased estimators of the true slope. The mean simulated two-point *z*ROC slope was equal to the ratio of the standard deviations of New and Old items when the two criteria were sampled in a range yielding false-alarm rates and remember false-alarm rates consistent with those in the literature. We conclude that the differences in variance and in central tendency between the two-point *z*ROC slopes in the R-K literature and the *z*ROC slopes in the old–new recognition literature are real differences.

Implications for the one-dimensional signal-detection model. Our analysis of two-point *z*ROC slopes shows that item-recognition and R-K data cannot be described by a single one-dimensional SDT model. Because remember judgments are assumed to be merely high-confidence old decisions, the same model must fit both the old–new data and the R-K data, and the slopes of the old–new *z*ROC curves and the two-point *z*ROCs derived from R-K data must be the same, either 1.0 (if the equal-variance model is correct) or about 0.8 (if an unequal-variance model is appropriate). In fact, however, the mean and variance of the two-point slopes observed in the old–new data and the two-point R-K data are substantially different. Consequently, these two-point ROC curves are not traditional ROCs, and the two points do not differ just in confidence. Contrary to both versions of the one-dimensional model, remember responses are not high-confidence old judgments, and old–new decisions and R-K judgments are not based on the same underlying familiarity dimension. R-K judgments must reflect some other form of evidence instead of, or in addition to, global memory strength.

The Sum–Difference Theory of REmembering And Knowing (STREAK)

We propose a new two-dimensional model of R-K judgments that is more faithful to Tulving's (1985) original idea that remember and know responses reflect the contributions of two phenomenally distinct forms of memory, global familiarity and specific recollection. However, our model retains the signal-detection advantages of the one-dimensional model, such as distinguishing accuracy from response bias and accounting for the level of confidence with which memories are reported. In the new model, memory strength varies not only on a global, overall strength dimension (as in the global matching models of memory, e.g., MINERVA, Hintzman, 1988; SAM, Gillund & Shiffrin, 1984) but also on a dimension that measures the specific strength of details associated with test items. Thus, the global strength of an item can be corroborated, or contradicted, by the specific strength of particular details.

The representation of specific evidence as a continuous variable is related to the idea of "differentiated" information in the source-monitoring framework (Johnson, Hashtroudi, & Lindsay, 1993; Mitchell & Johnson, 2000). In that view, memories retain the qualitative characteristics of encoding events, including perceptual information (such as color and sound) and reflective information (such as imagery and associations). In addition, the framework allows for the availability of these specific characteristics to vary from item to item, even among unstudied foils. Similarly, our model proposes that recollection is a graded form of evidence, like familiarity. Distributions of studied and unstudied items differ in their specific strength as well as global strength: Previously studied words are typically more familiar than unstudied words, and they also activate more orthographic and semantic information in memory.

The representation of the distributions of memory strengths of Old and New items is shown in Figure 4. Both studied and new items vary in their global memory strength or familiarity (along the *x*-axis) and in the strength of their more specific features or characteristics (along the *y*-axis). In the graphical representation of the one-dimensional model (see Figure 1), a second dimension indicates the likelihood of old and new memory strengths. In the STREAK model, analogously, the frequency of occurrence of all possible combinations of memory strengths could be indicated by a third dimension projecting directly out of the image. Instead, the New and Old distributions are represented by equal-probability contours that correspond to horizontal slices through three-dimensional, bivariate normal distributions. We assume that the global and specific strengths of studied items are more variable than those of New items, as indicated by the standard deviations of the two distributions in Figure 4.

To determine a response, the participant divides the space into regions using decision bounds, as in general recognition theory (Ashby & Townsend, 1986). Two linear bounds are used. An old–new decision bound divides the studied items from the New items, with a slope ($-d_y/d_x$) determined by the relative diagnosticity of global and specific information for making this discrimination:[4] The greater the mean memory strength of the Old distribution along either axis, the higher the criterion level of strength along that axis that may be required for an old response. With this old–new bound, old judgments are more likely as either the specific or global memory strength of an item increases. The two memory strengths combine additively (they *sum*) to determine whether an old or new response is appropriate; thus, a greater memory strength on one dimension can compensate for a lower strength on the other dimension to yield an old decision.

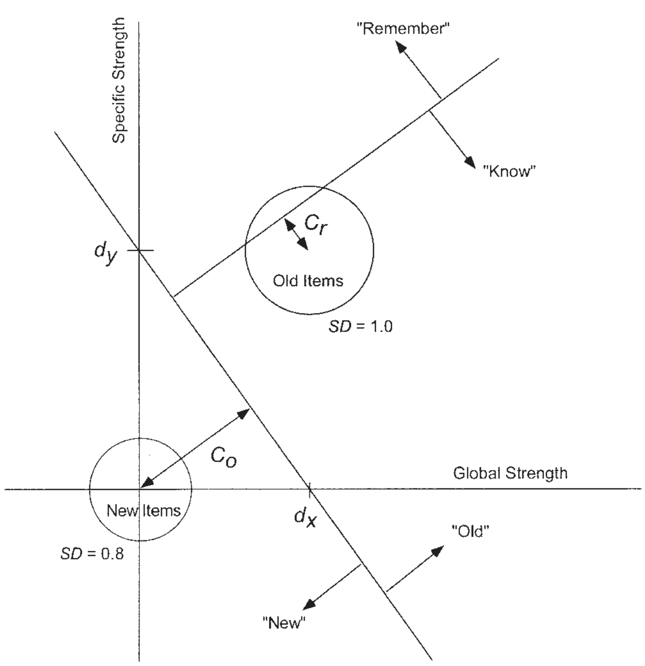

Figure 4. The sum–difference theory of remembering and knowing. Old and New items differ in both specific and global strength. One decision bound distinguishes old from new responses on the basis of a weighted sum of these axes, and a second bound distinguishes remember from know responses on the basis of a weighted difference. Circles represent equal-likelihood contours from bivariate distributions. d_x = diagnosticity of global information; d_y = diagnosticity of specific information; C_o = distance of the old–new decision bound from the mean of the New distribution; C_r = distance of the remember–know decision bound from the mean of the Old distribution.

An R-K decision bound divides items called "old" into those for which the specific memory strength is relatively great compared with global memory strength (remembered items) and those for which specific strength is relatively weak (known items). In other words, the decision is based on the *difference* between specific and global memory strengths: The more global strength an item has, the more specific strength is required to merit a remember judgment. The R-K bound is orthogonal to the old–new bound. Neither old–new nor R-K judgments are process pure (Jacoby, 1991) in this model: Both judgments are based on a joint assessment of global and specific information against a decision criterion that reflects their diagnosticities for the judgment. Thus, the weighted sum of global and specific strengths produces an old–new decision in the theory, and the weighted difference of strengths produces a remember or know judgment (sum–difference theory of remembering and knowing—STREAK).

An R-K experiment yields a 2 × 3 data matrix: Old and New items can each lead to remember, know, and new responses—with four degrees of freedom (because the remember, know, and new response rates must add to 1). Our analyses usually focus attention on the hit rate, the false-alarm rate, the *remember hit rate* (the probability of saying "remember" to an Old item), and the *remember false-alarm rate* (the probability of saying "remember" to a New item). To describe these four pieces of data, STREAK has four free parameters: the diagnosticity of global and specific information (d_x and d_y, respectively), the distance of the old–new decision bound from the mean of the New distribution (C_o), and the distance of the R-K bound from the mean of the Old distribution (C_r). All these parameters are measured in units of the standard deviation of the Old distribution. The New distribution has a smaller standard deviation *s* that is sometimes set to 0.80, the typical value in item-recognition experiments, rather than being estimated from data. Clearly, an R-K experiment provides just enough information to estimate the two-dimensional model's parameters; the equations for doing so are given in Appendix B.

Models of the R-K paradigm should account for two basic findings in the existing literature: performance in old–new item-recognition tasks (whether as the first part of the R-K judgment or in other contexts) and the slopes of two-point *z*ROCs constructed by treating remember and old responses as resulting from separate response criteria, as discussed earlier. We saw earlier that a single one-dimensional model cannot account for both kinds of data, and the first task of the new model is to show that it can surmount this hurdle.

[4] This decision bound does not in general intersect the points (0, d_y) and (d_x, 0) because the criterion placement can vary. However, the angle of the decision bound represented in Figure 4 is important to the model.

Item-recognition ROCs. STREAK accounts for the standard ROC findings in item-recognition (old–new) experiments. Although the model is two-dimensional, judgments of old versus new are one-dimensional, depending only on a weighted sum of these two values. The bivariate distributions can be projected onto a decision axis that extends from lower left to upper right in the space (perpendicular to the decision bound), and the criterion location is the point at which the decision bound passes through that axis. The projected New and Old distributions are univariate normal, with standard deviations of s and 1, respectively. By setting $s = 0.8$, we capture the observation in the literature that old–new zROCs have a slope of around 0.8 (Ratcliff et al., 1992). An item-recognition ROC can be generated by systematically varying the old–new criterion, C_o, from a conservative placement (toward the upper right in Figure 4) to a more liberal position (toward the lower left). Because the participant is providing a rating of old versus new and the curves plot the proportion of old responses, these item-recognition ROCs can also be termed *O-N rating/"old"* ROCs. The new terminology aids in comparing these curves with new types of ROCs to be introduced later.

Two-point zROC slopes. To evaluate the ability of the two-dimensional model to account for the two-point zROC curves, we varied C_o and C_r simultaneously and computed the predicted probabilities of remember and old responses to New and Old items. Normal transformations of these predicted probabilities allowed the calculation of predicted slopes of the two-point zROC from Equation 1.

The old–new criterion, C_o, was varied over 3 levels to produce roughly the same range of false-alarm rates as are typically found in the data: 0.5 (i.e., one half of a standard deviation above the mean of the New distribution), 1.0, and 1.5. To produce the same range of proportions of remember responses as are found in the data, we varied the R-K criterion, C_r, over 10 levels: from -0.2 (i.e., two tenths of a standard deviation to the left of the mean of the Old distribution) to 0.7 in steps of 0.1. In addition, d_x was varied between 0.5 and 2.0, and d_y was varied over the range 0.5–3.0, in steps of 0.1 units. The result was a very wide range of possible slopes, from about 0.1 to 400. The primary dependence of these slopes on the parameters of the model was that slopes of 1.0 or greater arose when d_y was greater than or equal to d_x.

Next, STREAK was fit to experimental data from each of the 373 conditions listed in Appendix A with Equations B1–B4 in Appendix B; because of the lack of an analytic solution, a grid-search method was used to find the parameter values. The grid search evaluated the fit of all possible combinations of parameter values in the following ranges: $0 < d_y < 2.5$, $0 < d_x < 2.5$, and $-2 < C_r < 2$, each varied in steps of 0.1 units. (C_o was computed directly from the observed false-alarm rate with Equation B1.) The parameter values that yielded the smallest sum of squared deviations (SSD) between the predicted and observed hit rates, remember hit rates, and remember false-alarm rates were selected as the best set for each experimental condition; these are shown in Appendix A. The mean SSD between model and data was 0.0004; the standard error was 0.00006; and the correlations of model-predicted and observed hit rates, remember hit rates, and remember false-alarm rates were all greater than .99. Thus, the grid search was successful in finding appropriate parameters.

Finally, we used the predicted hit and false-alarm rates to calculate the slopes of the two-point zROCs, using Equation 1. STREAK described the data well ($r = .914$, $p < .001$), and no systematic deviations between the predicted and observed data appeared. This correlation is lower than those between observed and predicted hit rates or false-alarm rates, mainly because even small deviations between the observed and predicted false-alarm rates can cause relatively large changes in the slope of the predicted two-point zROC. Nevertheless, the two-dimensional model can account for the problematic two-point R-K slopes, and it does so without contradicting the recognition slope data in the literature.

The cost of this success is the use of four parameters rather than three, and thus a saturated model. Although saturated models have no degrees of freedom remaining to test goodness of fit, they can be tested in three ways. First, parameter estimates should agree with values from other data. We have already seen a counterexample of this in evaluating the unequal-variance version of the one-dimensional model. If the ratio s of the Old and New standard deviations is allowed to be a free parameter, the model is saturated and "fits" perfectly, but the values of s obtained from two-point zROCs are in conflict with those obtained from ROCs in item recognition (see Figure 3). Second, an incorrect model may be forced to assume parameter values that are unrealistic for the situation being described. Several plausible modifications to the decision bounds, but not to the number of parameters, exhibit systematic over- or underprediction of the two-point R-K slopes. These alternative versions of the model are presented in the General Discussion. Third, the R-K paradigm can be expanded so as to increase the number of degrees of freedom; we describe this strategy next.

Some New ROC Curves

We have seen that ROC curves generated by confidence ratings have been an important tool in testing models of recognition memory. In this context, they are interpreted as revealing multiple adjustable criteria. STREAK implies two such manipulations because it has two decision bounds: Participants should be able to rate their memories not only on an old–new confidence scale (*O-N rating*) but also on a scale representing the degree to which their memory feels like remembering or knowing (*R-K rating*). It is also possible to examine how each type of criterion shift affects the probability of old responses as well as the probability of remember (and know) responses. Figure 5 shows the presumed effect of O-N ratings (see Figure 5A) and R-K ratings (see Figure 5B) on the placement of the decision bounds: Solid lines represent particular criterion placements used in a binary decision task; dashed lines represent additional decision bounds used in a rating task.

Table 1 defines the five types of ROCs that result from experiments in which a rating response of one type (O-N or R-K) is combined with a binary response of the other type. We refer to the various curves by both the rating required and the response evaluated—for example, O-N rating/"remember"—a terminological victory for precision over elegance. The R-K rating/"old" ROC is a single point because no changes in the old and new proportions are produced when the R-K bound is shifted. For R-K ratings, the remember and know curves are redundant, as the two responses add up to the (constant) proportion of old responses. For O-N ratings, two of the three curves are independent, and the third can be calculated from them. For example, the remember and know curves add to the (now variable) old curve.

The two-dimensional model's predictions about the shapes of these curves can be derived from the sketches in Figure 5. We

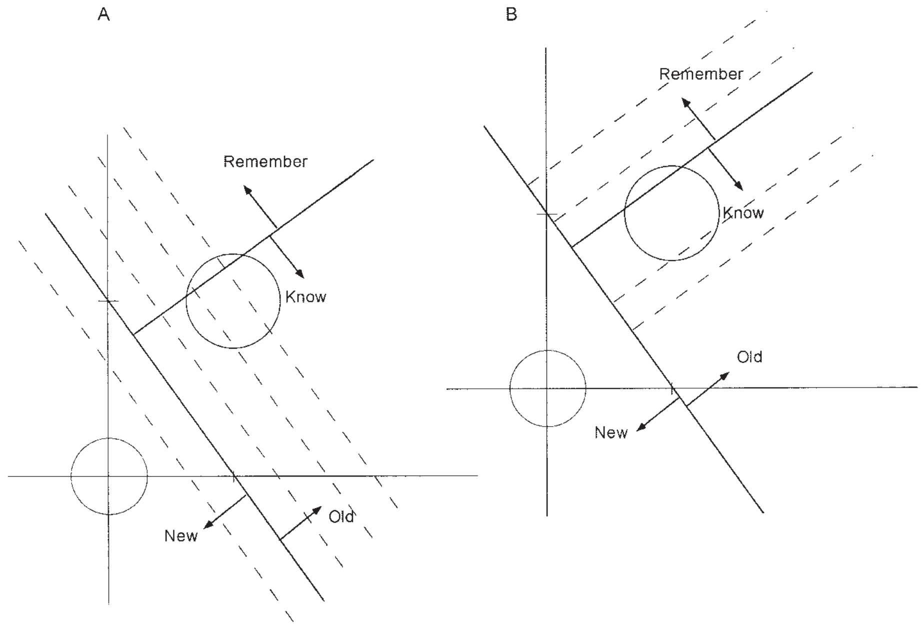

Figure 5. Schematic effects of rating instructions according to the sum–difference theory of remembering and knowing. In both panels, the solid lines represent memory strength axes and particular decision bounds; the dashed lines represent alternative decision bounds. In A, confidence ratings on old–new judgments correspond to changes in the placement of the old–new decision bound. In B, confidence ratings on remember–know judgments correspond to changes in the remember–know decision bound. Circles represent equal-likelihood contours for bivariate distributions.

consider them next, along with predictions based on the one-dimensional model. In addition, we examine the implications of the dual-process model (Yonelinas, 1994).

Predictions of the STREAK model. We have already described the shape of the O-N rating/"old" ROC; it is an unequal-variance ROC. Now, consider the two other ROCs from the O-N rating task, the remember and know curves. The O-N rating/"remember" ROC is generated by plotting the probability of a remember response to Old and New items as a function of old–new confidence. The O-N rating/"know" ROC, analogously, plots the probability of a know decision. Although they are not independent (because remember + know = old), it can be useful to consider these curves separately. Predicted remember curves are shown in Figure 6A for a particular R-K criterion; the curve was traced out by varying the old–new criterion, C_o, from −1.8 to 2.0. Because making a remember response is conditional on having made an old response, the probability of a remember response to both studied items and distractors is limited by the placement of the old–new criterion. These curves do not, therefore, rise to (1, 1). When the old–new criterion is relatively liberal, however, a remember curve can include higher remember hit and false-alarm rates than when the old–new criterion is stricter. In fact, the theoretical maximum point on each curve, which would occur if all old judgments were followed by remember responses, is determined by the true hit and false-alarm rates at each old–new criterion. The O-N rating/"know" curve for the same parameter values is also shown in Figure 6A.

Two different ROCs can be generated from R-K ratings, the remember and know curves. A predicted R-K/"remember" ROC curve generated with this approach is shown in Figure 6B for a particular old–new criterion. The curve was traced out by varying the R-K criterion, C_r, from −1.8 (i.e., very few remember responses) to 2.0 (i.e., very few know responses). The curve does not extend to (1, 1) because the probability of a remember response to both Old and New items is limited by the placement of the old–new criterion, just as in the O-N rating/"remember" and

Table 1
Types of ROC Curves That Can Be Generated Within the Sum–Difference Theory of Remembering and Knowing

Task	Response: Old	Remember	Know
O-N rating, R-K binary	O-N rating/"old"	O-N rating/"remember"	O-N rating/"know"
O-N binary, R-K rating		R-K rating/"remember"	R-K rating/"know"

Note. ROC = receiver operating characteristic; O-N = old–new; R-K = remember–know.

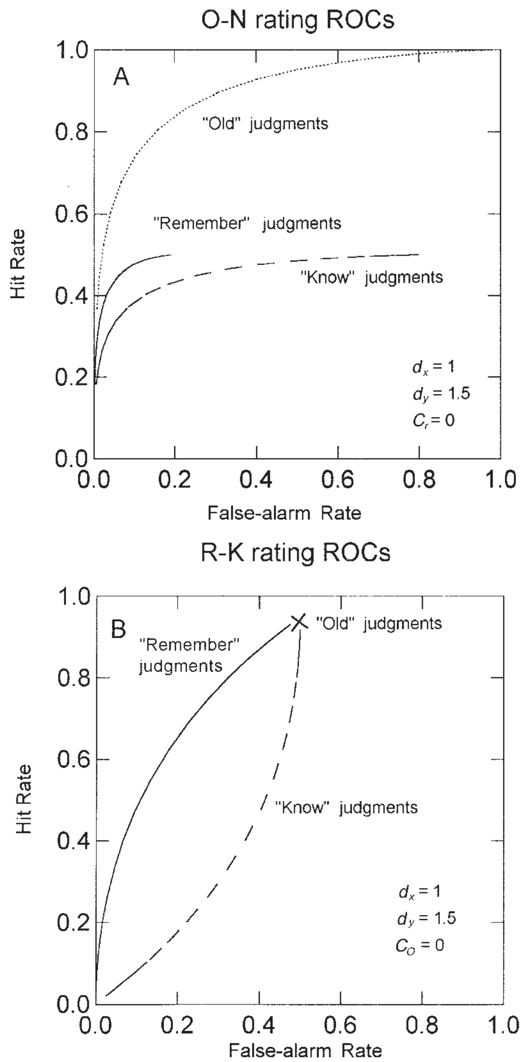

Figure 6. A: Old–new (O-N) rating/“remember,” “know,” and “old” receiver operating characteristic (ROC) curves predicted by the sum–difference theory of remembering and knowing (STREAK). Each curve traces out the effects of decreasing C_o. B: Remember–know (R-K) rating/“remember,” “know,” and “old” ROC curves predicted by STREAK. Each curve traces out the effects of decreasing C_r. d_x = diagnosticity of global information; d_y = diagnosticity of specific information; C_r = distance of the remember–know decision bound from the mean of the Old distribution; C_o = distance of the old–new decision bound from the mean of the New distribution.

“know” ROCs. Analogous R-K rating/“know” ROC curves can be generated by varying the R-K criterion, C_r, from 2.0 (very few know responses) to -1.8 (very few remember responses); an example is shown in Figure 6B.

Predictions of the one-dimensional model. In the one-dimensional model, remember responses result from familiarity values above a high remember criterion, and know responses result from those between that high criterion and a lower old–new criterion (see Figure 1). To our knowledge, no one has previously tested the one-dimensional model using ROC data; we assume that an observer produces O-N ratings by moving the old criterion while leaving the remember criterion fixed and R-K ratings by moving the remember criterion while leaving the old criterion fixed. This decision rule is parallel to our assumption about the two-dimensional model.

In an O-N rating experiment, the old criterion moves from the location of the remember criterion to the low end of the decision axis. The resulting old curve traces out the upper part of the standard old–new ROC (see Figure 7A), the lowest obtainable point being (P(“remember”|New), P(“remember”|Old)). The remember “curve” is that point alone because movement of the old–new criterion should have no influence on remember responses. The know response rate is the difference between the old rate and the (constant) remember rate; so, the know curve has the same form as the old curve but is shifted down from (P(“remember”|New), P(“remember”|Old)) to (0, 0).

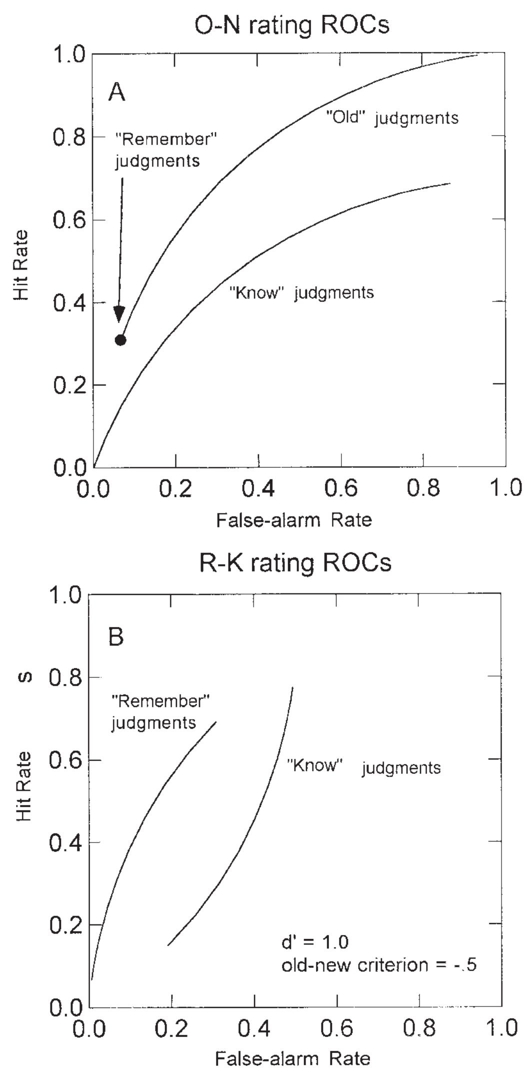

Figure 7. A: Old–new (O-N) rating/“remember,” “know,” and “old” receiver operating characteristic (ROC) curves predicted by the one-dimensional model. The remember curve is a single point. The curves trace out the effects of decreasing the old–new criterion from a value equal to the remember–know criterion. B: Remember–know (R-K) rating/“remember” and “know” ROC curves predicted by the one-dimensional model. The curves trace out the effect of decreasing the remember–know criterion to a value equal to the old–new criterion.

In an R-K rating experiment, the remember criterion moves from the high end of the decision axis to the location of the old criterion. The resulting remember curve traces out the lower part of the standard old–new ROC (see Figure 7B), the highest obtainable point being (P(“old”|New), P(“old”|Old)). (This curve is the complement of the O-N rating/“old” curve when the remember and old criteria are identical.) The know response rate is again the difference between the old and remember rates, but now the old rate is a constant; so, the know curve has the inverse shape to that of the remember curve.[5]

Implications of the dual-process model. An alternative model of recognition memory, which has occasionally been applied to R-K judgments (Yonelinas, 2001; Yonelinas, Dobbins, Szymanski, Dhaliwal, & King, 1996), is the dual-process recognition model proposed by Yonelinas (1994). According to this model, recognition decisions are based on a familiarity process (an equal-variance SDT model like that in Figure 1) and a separate recollective process. Studied items are called “old” when they can be recollected or when they are not recollected but exceed a criterion level of familiarity. Because recollection is modeled as a high-threshold process, only test items that were actually studied can be recollected; distractors can be called “old” only on the basis of familiarity. The dual-process model predicts old–new ROC curves that are generally similar to unequal-variance one-dimensional SDT curves in probability space (see Yonelinas, 1994, 1997): They have a nonzero y-intercept that indicates the proportion of items that are recollected and a curvilinear shape that increases smoothly to the point (1, 1).

Like the one-dimensional model, the dual-process model predicts remember and know ROC curves that are quite distinct from those predicted by the two-dimensional model. In the dual-process model, the recollection parameter can be estimated with the remember hit rate minus the remember false-alarm rate (Yonelinas et al., 1996; Yonelinas & Jacoby, 1995). The high-threshold nature of the recollective process implies that varying familiarity-based decision criteria cannot change the proportion of items that are remembered; so, any New items that are remembered must be treated as measuring nonmemorial noise. Alternatively, remember responses may be treated simply as “high confidence recognition responses” (Yonelinas, 2001, p. 367).[6] Regardless of which approach is taken, the O-N rating/“remember” curve is a single point in the dual-process model, (P(“remember”|New), P(“remember”|Old)), as shown in Figure 8. That point does not vary as a function of the old–new decision criterion. Yonelinas has reported some data consistent with this hypothesis: Remember responses occurred almost exclusively after highest confidence old judgments, yielding a single remember hit rate and a single remember false-alarm rate (Yonelinas, 2001; Yonelinas et al., 1996).

The know false-alarm rate is not truly limited by the remember rate because no false alarms can result from remembering the item. It is reduced somewhat, however, by the nonmemorial noise that may generate a few remember false alarms. As a consequence, the O-N rating/“know” ROC covers almost the entire range of possible false-alarm rates (i.e., from 0 to 1 minus the remember false-alarm rate). The know hit rate is restricted by the items that are recollected (and therefore remembered rather than known); so, the curve does not approach (1, 1). Figure 8 provides example O-N rating/“know” ROC curves generated from the dual-process model. The dual-process model has not been developed so as to predict R-K rating curves; in fact, the threshold nature of the recollection process seems to imply that they cannot be produced experimentally. We comment more on this in the General Discussion.

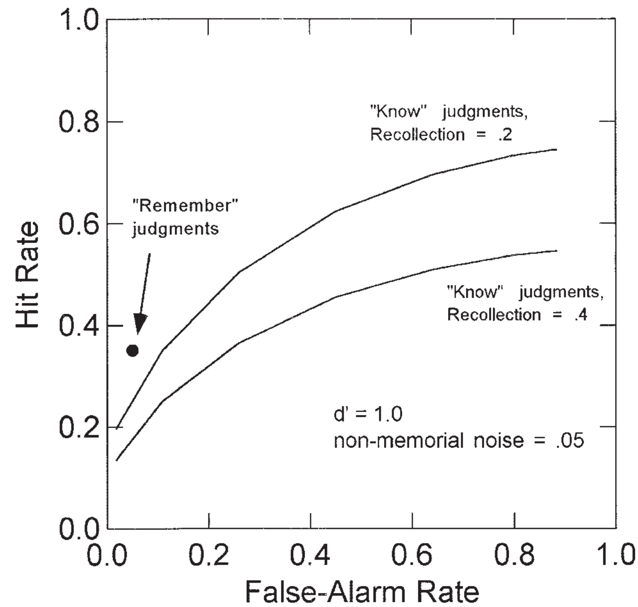

Figure 8. The old–new rating/“remember” receiver operating characteristic (ROC) curve predicted by the dual-process model is a single point because remember responses are given only to highest confidence responses. Two old–new rating/“know” ROC curves, for different values of R (the probability that a studied item is recollected), each trace out the effect of decreasing the criterion.

All of the models make predictions about the shapes of old, remember, and know ROCs generated by varying the placement of the old–new criterion while the R-K criterion (or threshold) is fixed. ROC data of this type were collected in Experiment 1.

Experiment 1

In Experiment 1, participants studied a list of words and then rated their confidence that test words were old. After each old

[5] The ROCs in Figure 7 assume the simplest response strategies, in which only one criterion is moved. No matter which criterion is shifted, the hit and false-alarm rates change in the same manner; thus, the O-N/“old” and “remember” curves in Figure 7A and the R-K/“remember” curve in Figure 7B are segments of the same normal–normal ROC. A more complex decision rule allows both criteria to move together, under either O-N or R-K ratings. With this strategy, old and remember responses still trace out parts of the same ROC (whether the distance between the criteria is fixed or not), but the prediction of a single point for the O-N rating/“remember” curve no longer holds.

[6] It is possible for the remember responses to Old items to be distributed over a range of confidence levels in the dual-process model, although we do not believe that this extension to the model has been formally developed. If one assumes that the remember responses to New items are similarly distributed, however, the recollective component of the model loses its high-threshold nature, which has been important for fitting the linear ROCs that are observed under certain limited conditions (e.g., Rotello et al., 2000; Yonelinas, 1997).

response, they also reported whether they remembered or knew that the word had been studied. Thus, Experiment 1 allowed us to generate three kinds of ROC curves: O-N rating/“old” (a standard old–new ROC), O-N rating/“remember”, and O-N rating/“know.” Two versions, Experiments 1A and 1B, were conducted. The differences between them are minor and are described below.

Method

Participants. Fourteen University of Massachusetts undergraduates participated in Experiment 1A for partial course credit; data were discarded for 1 additional participant who confused the remember and know responses. Fifteen volunteers from the same pool participated in Experiment 1B; data were discarded for 1 additional participant who confused the remember and know responses and for another who exhibited chance-level recognition performance.

Stimuli. We selected 120 nouns with one or two syllables from the Medical Research Council Psycholinguistic Database (Coltheart, 1981). These words were divided into two lists of 60 that were closely matched for number of syllables ($M = 1.5$), number of letters ($M = 5.2$), and frequency of occurrence (Kučera & Francis, 1967; $M = 157$, $SD = 112$). An additional 14 words were drawn from the same pool to serve as practice, primacy, and recency items. The words were displayed in the center of a 17-in. computer monitor in a plain black font against a white background. Response alternatives for the memory test were mapped onto an array of eight labeled keys, including the numbers *1–6*, *z*, and */*.

Procedure. The procedure consisted of a study phase followed by a practice phase and a final test phase. In the study phase, all 60 words from one list were combined with 4 practice words and presented in random order, preceded by 3 primacy words and followed by 3 recency words. The lists were counterbalanced across participants. In Experiment 1A, each word appeared for 1,000 ms with a 500-ms interstimulus interval; in Experiment 1B, each word appeared for 1,250 ms with a 750-ms interval. Participants were instructed simply to read each word carefully in preparation for the upcoming memory test.

In the practice phase, the 4 practice words from the study phase were combined with the 4 unstudied practice words and presented as memory probes. Similarly, the test phase included the 60 words from the study phase (excluding primary and recency words) combined with the list of 60 unstudied words, presented as memory probes in random order. The practice and test phases followed the same procedure, except that participants were encouraged to ask questions during the practice phase.

For each probe, participants performed a one- or two-part memory judgment. First, they judged whether they had previously studied the probe, rating their confidence on a 6-point scale. In Experiment 1A, the points of the scale were labeled (1) *certain yes*, (2) *probably yes*, (3) *maybe yes*, (4) *maybe no*, (5) *probably no*, and (6) *certain no*. The points of the scale in Experiment 1B were more sparsely labeled: (1) *certain yes*, (5) *maybe yes*, and (6) *certain no*.[7] Whenever participants called a probe “old” (by responding 1–3 in Experiment 1A or 1–5 in Experiment 1B), they were also prompted to make a binary judgment that described their feeling of recognition as either remembering or knowing. Thus, R-K responses were collected for three levels of recognition confidence in Experiment 1A and for five levels in Experiment 1B.

Participants were instructed to base their R-K judgments on the descriptions of remembering and knowing provided by Rajaram (1993). Remembering was described as “the ability to become consciously aware again of some aspect or aspects of what happened or what was experienced at the time the word was presented (e.g., aspects of the physical appearance of the word, or of something that happened in the room, or of what you were thinking and doing at the time)” (Rajaram, 1993, p. 102). In contrast, knowing was described as the feeling that “you recognize that the word was in the study list but you cannot consciously recollect anything about its actual occurrence or what happened or what was experienced at the time of its occurrence” (Rajaram, 1993, p. 102). In addition to these standard descriptions, the instructions pointed out that although remembering and knowing are different feelings of recognition, both may vary in confidence. Finally, the instructions included an example about recognizing people at the grocery store: “Sometimes you remember the reason you recognize them, such as the circumstances of previous encounters, but other times they just seem familiar, so that you know you recognize them even though you cannot say why” (Rajaram, 1993, p. 102).

The practice and test phases were self-paced; most participants completed all 120 test trials in 10–15 min.

Results

O-N rating/“old” ROC data. Figure 9 presents the ROCs from both Experiments 1A and 1B. To fit the model to the data, we set the overall hit and false-alarm rates to the probabilities of old–new ratings of 1–3 and 4–6, respectively, and the remember hit and false-alarm rates to the proportions of Old and New items given a remember response. The parameter values for STREAK were then found from Equations B1–B4 in Appendix B, with s set equal to the slope of the observed zROC in each experiment. We implemented these equations in Microsoft Excel’s Solver module, which uses a simplex method to find parameter values.[8] The resulting values, shown in Table 2, agree with those obtained by using the older grid-search method described earlier but have greater precision.

According to both the one- and two-dimensional models, the R-K criterion is not involved in old–new confidence ratings; so, the predicted ROC is either an equal- or unequal-variance SDT curve. Maximum likelihood estimation (implemented by SYSTAT’s signal-detection module) revealed that the slope of the zROC is 0.81 in Experiment 1A and 0.91 in Experiment 1B; so, the unequal-variance SDT model provides a better account of these data than the equal-variance model.

O-N rating/“remember” ROC data. Remember responses to studied words plotted against those to new words, as a function of old–new confidence level (from “certain” old to “maybe” old), are shown as the points in Figure 10 (Experiment 1A in 10A and Experiment 1B in 10B). The solid and dotted curves in the figure were generated by the unequal- and equal-variance versions of the one-dimensional model. These curves are simply the lower segments of the old–new ROC; so, the unequal-variance fit is better. The dashed curve, generated with STREAK (cf. Figure 5), provides the best description of the data. The parameter values were the same as those used to generate the O-N rating/“old” ROC (see Figure 9).

O-N rating/“know” ROC data. The O-N rating/“know” ROC curves are shown in Figure 10C for Experiment 1A and in Figure 10D for Experiment 1B. The theoretical curves in each figure are again model generated, with the same parameters as for the other ROCs.

Discussion

The results of Experiment 1 are generally consistent with the two-dimensional sum–difference model of remembering and

[7] Actually, the scale in Experiment 1B was labeled in the opposite direction, with strong old decisions being given a rating of 6. For clarity, we describe the scales as being ordered in the same way.

[8] Copies of the routine are available on request from Caren M. Rotello.

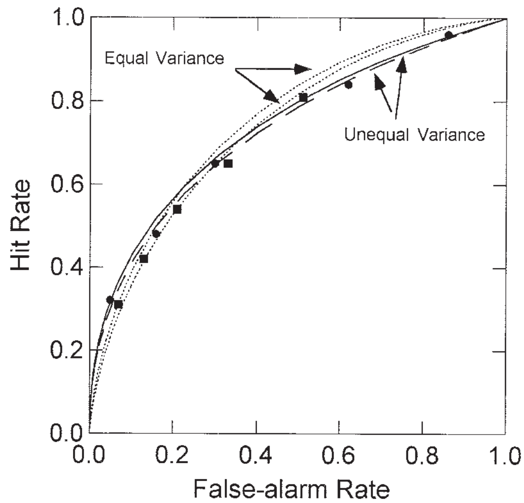

Figure 9. Old–new rating/“old” receiver operating characteristic (ROC) data from Experiments 1A (circles) and 1B (squares). The superimposed functions were generated with the unequal-variance one-dimensional model (dashed and solid lines) and with the equal-variance model (dotted lines). The predictions of the sum–difference theory of remembering and knowing are identical to those of the unequal-variance model.

knowing and inconsistent with aspects of both the one-dimensional and dual-process models of recognition. Specifically, the O-N rating/“old” ROCs, with slopes of 0.81 and 0.91 in Experiments 1A and 1B, were well-fit by the one-dimensional unequal-variance SDT model (but not by its equal-variance cousin) and by the two-dimensional model. The present results are consistent with the prior literature on the slope of the *z*ROC in item-recognition tasks (e.g., Ratcliff et al., 1992). Although it succeeds with item recognition, the unequal-variance one-dimensional model is inconsistent with the two-point R-K *z*ROC slopes, which were 0.85 and 1.03 in Experiments 1A and 1B, respectively. The dual-process model can fit the O-N rating/“old” ROC data, but only STREAK can describe both item-recognition ROCs and two-point *z*ROC slopes.

The O-N rating/“remember” ROCs also support the two-dimensional model over its competitors. The remember curves are truly curves, including several false-alarm rates rather than the single point predicted by both versions of the one-dimensional model and by the dual-process model (see Figures 7 and 8). Moreover, the remember ROC curves generated with the two-dimensional model fit the data well. That several distinct false-alarm rates are associated with remember responses is inconsistent with data reported by Yonelinas et al. (1996, Experiment 2; Yonelinas, 2001, Experiment 1). Participants in those experiments provided remember responses almost exclusively following highest confidence old decisions; so, a single remember false-alarm rate (and a single remember hit rate) was observed. One primary difference between Yonelinas’s studies and ours is that his participants were told exactly what detail they needed to recollect to justify a remember decision, whereas ours were given traditional R-K instructions in which any recollected detail was sufficient for a remember response. We recently manipulated this factor between subjects in a single experiment (Rotello, Macmillan, Reeder, & Wong, 2004, Experiment 3). Participants heard words in either a male or female voice and were asked to say “remember” either (a) whenever they could recollect any detail of the experience of studying the test word or (b) only when they could report who said the word to the experimenter. The standard definition of a remember response resulted in remember ROC curves that looked much like those in Figure 10. In contrast, the narrow definition of remembering, like that used in Yonelinas’s experiments, resulted in single-point remember ROCs: Participants made remember responses exclusively after highest confidence old decisions. The two-dimensional model fit both types of data.

Finally, the O-N rating/“know” ROC data show an advantage for the unequal-variance one-dimensional model, with the observed data being somewhat more linear than the two-dimensional model predicts. This discrepancy may be accounted for by the inclusion of some guess responses in the know judgments (e.g., Gardiner, Richardson-Klavehn, & Ramponi, 1997), which would tend to shift the know ROC data points toward chance levels. We consider an alternative account of this pattern of data—that the two decision criteria are not fully independent—in the General Discussion.

Experiment 2

A striking implication of the STREAK model is that participants can make graded R-K decisions: They can rate their relative degree of remembering and knowing, just as they rate their confidence that an item is old. We tested this prediction in Experiment 2. Participants studied a list of words and then decided whether test words had been studied or not. After each old response, they rated their degree of remembering or knowing. Thus, Experiment 2 allowed us to generate R-K rating/“remember” and “know” ROC curves. The two versions of Experiment 2, 2A and 2B, exactly parallel Experiments 1A and 1B.

Method

Experiments 2A and 2B were conducted concurrently with Experiments 1A and 1B, respectively.

Participants. Fourteen University of Massachusetts undergraduates volunteered to participate in Experiment 2A for partial course credit. Fourteen volunteers from the same pool participated in Experiment 2B; data were discarded for 1 additional participant who exhibited chance-level recognition performance.

Stimuli. The word lists, presentations, and response keys in Experiment 2 were identical to those in Experiment 1.

Table 2
Best Fitting Parameter Estimates From Experiments 1 and 2

Experiment	Best fitting parameter estimates				
	s	d_x	d_y	C_r	C_o
1A	0.81	0.45	0.93	0.21	0.42
1B	0.91	0.49	0.87	−0.08	−0.02
2A	0.80	0.52	1.20	0.38	0.38
2B	0.80	0.51	0.94	0.39	0.49

Note. For Experiments 1A and 1B, s is the slope of the old–new rating normal–normal receiver operating characteristic. For Experiments 2A and 2B, s was fixed at 0.80. d_x = diagnosticity of global information; d_y = diagnosticity of specific information; C_r = distance of the remember–know bound from the mean of the Old distribution; C_o = distance of the old–new decision bound from the mean of the New distribution.

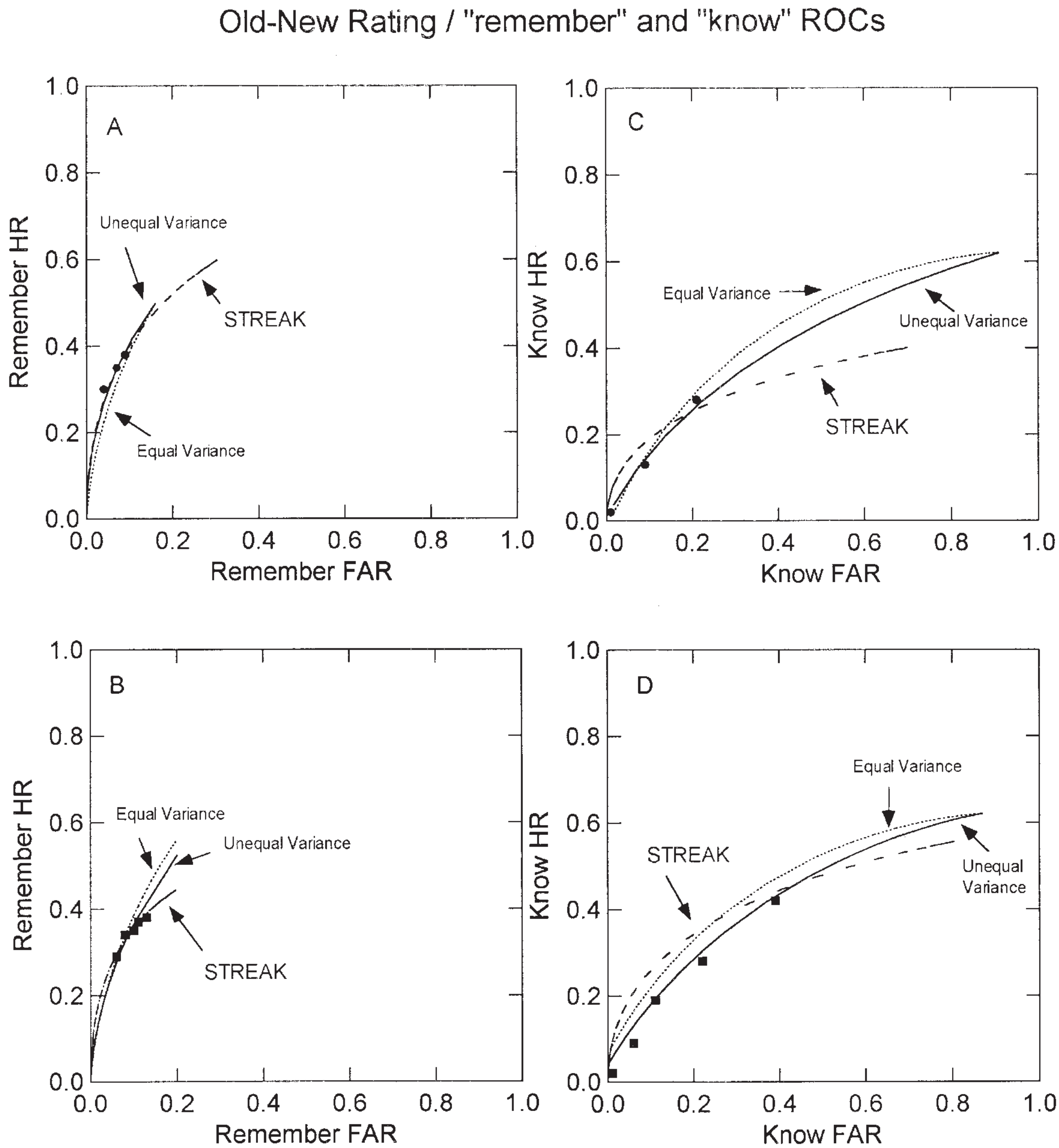

Figure 10. Old–new rating/"remember" and "know" receiver operating characteristic (ROC) data from Experiments 1A (A and C) and 1B (B and D). The superimposed curves were generated with the sum–difference theory of remembering and knowing (STREAK; dashed lines) and both the equal- (dotted lines) and unequal-variance (solid lines) versions of the one-dimensional model. HR = hit rate; FAR = false-alarm rate.

Procedure. Experiments 2A and 2B included study phases identical to those in Experiments 1A and 1B, respectively. The practice and test phases were similar to those in Experiment 1, except for an important modification to the formats of the recognition and R-K judgments. Participants first judged whether they had previously studied each probe, but only the binary response alternatives "yes" and "no" were available, in contrast to the confidence-rating scale in Experiment 1. Whenever participants indicated that they did recognize a probe, they were also prompted to make an R-K rating.

As in Experiment 1, the R-K instructions explained that although remembering and knowing are different feelings of recognition, both can vary in confidence. Unlike the binary R-K judgment in Experiment 1, a 6-point rating scale was provided so that participants could characterize their remember and know feelings over a range of intensities. In Experiment 2A, the points of the scale were labeled (1) *remember lots of details,* (2) *remember some details,* (3) *remember few details,* (4) *weak feeling of knowing,* (5) *moderate feeling of knowing,* and (6) *strong feeling of knowing.* In Experiment 2B, the scale was anchored only at the endpoints of the scale: (1) *remember specific aspects of the experience* and (6) *know it feels very familiar, but nothing specific.* Participants were free to determine their own subjective basis for responding on the scale in Experiment 2B, as long as the two extreme ratings (1 and 6) corresponded to strong feelings of remembering or knowing as described by Rajaram's (1993) instructions. In other words, the intermediate responses could correspond to combined feelings of remembering and knowing or to weaker feelings of either type relative to the extreme ratings.

As in Experiment 1, the practice and test phases were self-paced, with most participants completing the 120 test trials in 10–15 min.

Results

R-K rating/"remember" ROC data. Remember responses to studied words are plotted against those to new words, as a function of R-K rating level (summing from "lots of details" to "strong feeling of knowing"), in Figure 11 (Experiment 2A in Figure 11A and 2B in Figure 11B). To compare the data to STREAK's predictions required the estimation of parameter values for this condition. As in Experiment 1, we found the overall hit and false-alarm rates as well as the remember hit and false-alarm rates (a remember response was defined as a rating of 1, 2, or 3 on the R-K scale in Experiment 2A). These probabilities were used with Equations B1–B4 in Appendix B, with $s = 0.8$, to produce the parameter estimates shown in Table 2. Also as in Experiment 1, we used Microsoft Excel's Solver module to obtain these estimates, which were in agreement with those obtained using the older grid-search method. The parameter values in Table 2 were used to generate the dashed remember ROC curves that are superimposed on the data in Figure 11. As can be seen, the two-dimensional model fits the data well.

We also fit the one-dimensional model to the data from each experiment, in both the equal- and unequal-variance forms. The data in Experiment 2A show the largest differences in the models'

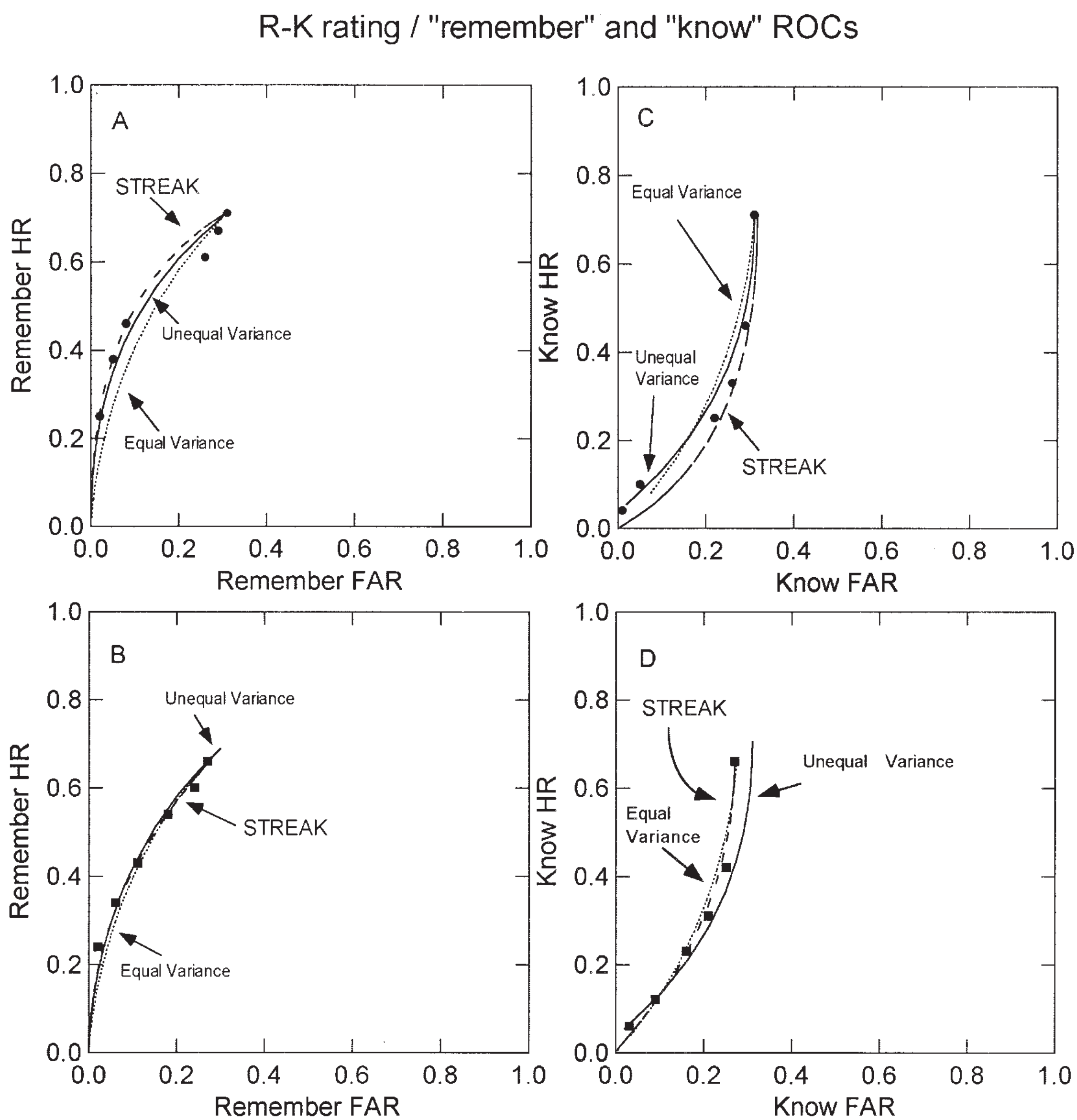

Figure 11. Remember–know (R-K) rating/"remember" and "know" receiver operating characteristic (ROC) data from Experiments 2A (A and C) and 2B (B and D). The superimposed curves were generated with the sum–difference theory of remembering and knowing (STREAK; dashed lines) and both the equal- (dotted lines) and unequal-variance (solid lines) versions of the one-dimensional model. HR = hit rate; FAR = false-alarm rate.

predictions, revealing an advantage for the unequal-variance one-dimensional model over the equal-variance version. (A similar but much smaller effect is present in Experiment 2B.)

R-K rating/"know" ROC data. The know responses to studied words are plotted against those to new words, as a function of R-K rating level (summing from "strong feeling of knowing" to "lots of details") in Figures 11C (Experiment 2A) and 11D (Experiment 2B). Superimposed on the data are STREAK's R-K rating/"know" ROC curves, generated with the same parameter values as the remember ROC curves in Figures 11A and 11B. The data are well-described by the model (as they must be because the remember data were well-fit by the model and the know ROC data simply cumulate the same responses from the other end of the R-K continuum).

According to the one-dimensional model, Experiment 2 allows the participant seven levels of confidence: In Experiment 2A, these are remember (three levels), know (three levels), and new; in Experiment 2B, the partition between remember and know is left to the participant, but the same number of levels is available. If the data are cumulated from highest to lowest confidence, the one-dimensional model predicts a normal–normal ROC with a z slope of about 0.8. The two versions of the one-dimensional model again fit less well than the two-dimensional model, especially in Experiment 2A (see Figure 11C).

Discussion

The two-dimensional model fits the R-K rating data very well, and the unequal-variance one-dimensional model also provides a good fit. Consistent with all SDT models, participants are able to use R-K ratings to reveal graded information about the basis of their item-recognition response. These results are a challenge to the dual-process model, in which remembering is a threshold process that is quite distinct from the familiarity-based assessment of knowing. The very idea that remembering and knowing are two ends of a continuum seems contrary to the dual-process model. The best that the dual-process model can do to fit the data from Experiment 2 is to assume that the same remember and know ROC curves would emerge as when the O-N rating task was used (i.e., as in Experiment 1). A single-point remember ROC and a concave-down know ROC would be predicted; neither of these predictions is supported by the data.

One might speculate that our instructions to spread out the remember responses over a range of ratings may have forced the curved, multipoint remember ROC to emerge in the data. However, the implications of whether participants can be instructed to obey a model are asymmetric. If there really are multiple cognitive states (several levels of remembering) and participants are encouraged to believe that R-K is a dichotomy, then it would not be surprising if they collapsed their responses into a single category. If, however, there is just one remember state, then participants who are trying to follow our instructions can do so only by choosing ratings randomly, and the data would reflect the representation accurately (e.g., linear ROCs). Our data were curvilinear, a result that is inconsistent with a threshold model under either set of instructions.

General Discussion

In the R-K paradigm, participants provide two pieces of information about their memory for each test item, an old–new judgment and an R-K judgment. Our STREAK model proposes that these judgments depend on the strengths of global familiarity and specific information about the item, with old–new judgments being a weighted sum and R-K judgments a weighted difference of these two strength values. The two dimensions of the model are both continuous (rather than discrete or thresholdlike), and decisions are made in a signal-detection manner.

Using ROC curves, we tested this two-dimensional description of R-K experiments against four sets of data:

1. old–new judgments obtained in rating experiments,
2. remember and old response rates plotted as two-point curves in ROC space,
3. R-K judgments obtained in old–new rating experiments, and
4. R-K judgments obtained in R-K rating experiments.

Data of all types were collected as part of the present investigation, and Data Sets 1 and 2 also include a substantial body of previous work. The model accounted for all types of data successfully. A summary of the database is provided in Table 3.

Comparison With Other Models of R-K Judgments

The primary competition for our account comes from the one-dimensional model (Donaldson, 1996; Hirshman & Master, 1997), which proposes that remember and know judgments reflect different degrees of strength on a unitary axis. The ability of both the equal- and unequal-variance forms of this model to describe the data is also summarized in Table 3. The equal-variance model does not handle item recognition (Data Set 1) very well because the slope of the zROC is expected to be 1.0 rather than in the neighborhood of 0.8, as found in the data. It does explain the two-point R-K ROC data (Data Set 2), but that is inconsistent with the observed O-N rating/"remember" ROCs (Data Set 3). The unequal-variance version is consistent with the item-recognition slope data (Data Set 1) but fails to describe the two-point R-K ROC data accurately (Data Set 2). In its simplest form, it is inconsistent with the observed old–new rating data (Data Set 3) because the remember curve is more than a single point, but it is able to describe R-K rating ROCs (Data Set 4).

We also fit a saturated version of the model in which the variance ratio was free to vary. It failed to fit the item-recognition data (Data Set 1) because the slope of the zROC was found to be at least 1. It failed to fit the two-point R-K ROC slopes observed in these experiments (Data Set 2) and is inconsistent with the observed O-N rating data (Data Set 3). In summary, the one-dimensional model cannot account for the entirety of the relevant data unless it is allowed to fluctuate in its form, sometimes having equal-variance distributions and other times having unequal-variance distributions.

The dual-process model provides a good account of old–new judgments from rating experiments (Data Set 1) but has not been developed in enough detail to make clear predictions about all aspects of the other data sets. (For example, two-point ROC slopes are difficult to predict because remember false alarms are not expected to occur and z slopes are undefined if one of the points has a false-alarm rate of 0.) We therefore do not attempt to include

Table 3
Summary of the Data Sets Against Which the Models Were Tested and the Success of Each Model

Data set and type of ROC	Source of data	Finding[a]	Predictions or fit of models[a]: One-dimensional, equal variance	One-dimensional, unequal variance	STREAK (saturated)
1					
O-N/“old” (conventional item-recognition) slope	Literature	0.77 ± 0.01[b]	1.0	**≈0.8**	**≈0.8**
	Expt 1A, 1B	0.81, 0.91[c]			
2					
Two-point slope	Meta-analysis	1.01 ± 0.02[b]	**1.0**	≈0.8	**≈1.0**
	Expt 1A, 1B	0.85, 1.03[c]			
	Expt 2A, 2B	0.93, 1.26[c]			
3					
O-N/“remember”	Expt 1	ROC is multipoint	Single-point prediction is disconfirmed; adequate fit if R-K criterion allowed to vary		**Good fit**
O-N/“know”	Expt 1	ROC is concave downward	Adequate fit	**Good fit**	Adequate fit
4					
R-K/“remember”	Expt 2	ROC is multipoint	Adequate fit	**Good fit**	**Good fit**
R-K/“know”	Expt 2	ROC is concave upward	**Good fit**	**Good fit**	**Good fit**

Note. Bold entries indicate good correspondence between the model and the data. ROC = receiver operating characteristic; STREAK = sum–difference theory of remembering and knowing; O-N = old–new; Expt = Experiment; R-K = remember–know.
[a] Numerical values are *z*ROC slopes. [b] $M \pm SE$. [c] M for each experiment, respectively.

it in Table 3. We have argued, however, that the simplest interpretation of the model is inconsistent with most of the data in Data Sets 3 and 4 (O-N and R-K rating remember and know ROCs).

Each of these models makes a strong but apparently incorrect assumption. In the case of the one-dimensional model, the fault lies in the consideration of only one type of memory strength. This strong counterintuitive postulate has made the model an attractive target for criticism, but it has two considerable virtues. First, the model is psychologically simple, and it is important to reject simple explanations before advancing to complex ones. Second, the model is explicit, permitting straightforward quantitative evaluation.

The dual-process model’s strong assumption is that remember responses have a threshold character. The most obvious hurdles facing this hypothesis are the O-N rating/“remember” ROCs (see also Rotello et al., 2004)—because a single-point ROC is predicted yet rarely observed—and the R-K rating task—because remembering and knowing are based on entirely different processes in the model. Although there is some limited but very clear evidence for threshold representations in recognition memory (e.g., Rotello, Macmillan, & Van Tassel, 2000; Yonelinas, 1997), the data are more typically well-fit by an unequal-variance SDT model of recognition as well as by the dual-process model.

Alternative Versions of STREAK

We have described the two-dimensional model as having linear old–new and R-K decision bounds that depend on both the global and specific memory strength axes. Other possible bounds that may seem at least equally natural are considered next. The particular bounds we postulate in our model can be justified by plausibility arguments, but they were chosen simply because the existing data are better fit by these particular bounds than by others. In this respect, STREAK is entirely data driven.

Likelihood-ratio decision bounds. The optimal decision rule uses a bound defined by likelihood ratio: Whenever the likelihood ratio that an item comes from the Old item distribution rather than from the New distribution is at least, say, 1.0 (at least even odds that it is old rather than new), the item is called “old.” If the standard deviation of the New distribution does not equal that of the Old distribution, the resulting bound is a curve rather than a straight line in the decision space. Various likelihood ratios can be chosen to require a greater or lesser probability that the test probe is actually an old item, and thus an old–new ROC curve can be generated.

We assumed that the standard deviation of the Old distribution was 1.0 and the standard deviation of the New distribution was 0.8, as in our previous simulations of the model. We set $d_x = 1.0$ and $d_y = 1.5$ (typical of the values found when the model is fitted to the data in Appendix A) and tested old–new likelihood ratios of 0.5, 1.0, and 1.5. The likelihood-ratio idea does not apply to the R-K criterion; so, we added a linear R-K bound with one of three different slopes: (a) d_x/d_y, as in the model shown in Figure 4; (b) a shallower slope of $0.5 \cdot (d_x/d_y)$ that is consistent with the idea that higher confidence old judgments and remember judgments may be correlated (see discussion below), and (c) a steeper slope of $1.5 \cdot (d_x/d_y)$ for completeness. Monte Carlo simulations with 10,000 trials each were conducted to produce estimated hit and false-alarm rates and remember hit and false-alarm rates. We then repeated these simulations with R-K decision bound slopes of $0.5 \cdot (d_y/d_x)$, d_y/d_x, and $1.5 \cdot (d_y/d_x)$. In all cases, the resulting R-K rating/“remember” and “know” curves had the wrong shape compared with our data and/or the predicted slopes of the two-point *z*ROCs were too steep.

Optimal linear decision bounds. Even if we limit the model to linear bounds, the bounds assumed by the two-dimensional STREAK model do not have optimal slope. The best choice is a decision bound that is perpendicular to a decision axis that intersects the means of the Old and New distributions. Because the slope of this axis is d_y/d_x, the optimal slope of the decision bound is $-d_x/d_y$. We implemented the model using this bound together with a variety of R-K bounds, but all failed on two counts: The two-point *z*ROC slopes were greater than 1.0, and the R-K rating/"remember" and "know" ROCs had incorrect form.

The bounds we have used do have undesirable properties if either d_x or d_y is extremely large. For a given value of d_x, as d_y increases toward infinity, the slope of the old–new decision bound (i.e., $-d_y/d_x$) approaches negative infinity, meaning that the old–new decision is based almost exclusively on information from the global memory strength axis. This is contrary to what one would expect: When the Old and New distributions are infinitely far apart on the specific memory strength axis, it would be sensible to use that information to the exclusion of global strength. An analogous problem emerges as d_x approaches infinity. Bounds with optimal slope do not have this problem. But, the nonoptimal bounds that participants apparently prefer do have at least one intuitive advantage: Criterion locations on both axes tend to increase with sensitivity, a strategy that limits bias. Also, in fitting the model to data, we did not find cases in which d_y was much larger than d_x, or vice versa, and dealing with the mathematical boundary conditions is arguably premature. The finding of nonoptimal decision making is of some interest but is not especially startling.

Process-pure decision bounds. An obvious alternative set of decision bounds is an R-K criterion that makes use of memory strength only on the specific-strength dimension, and a know bound that makes use of memory strength only on the global dimension. In such a model, illustrated in Figure 12, participants make remember judgments whenever the specific information about a test item exceeds a remember criterion, regardless of the global memory strength of that item. Failing that, know judgments are made whenever the global memory strength of the item exceeds the know criterion. (Old responses occur whenever the global familiarity exceeds the know criterion or the specific familiarity exceeds the remember criterion.) These decision criteria imply that remember and know judgments are process pure: Remember judgments involve only specific memory strength, and know judgments depend only on global familiarity in the absence of specific strength.

This structure of remember–know–new decision criteria has been implemented in a recent model of the development of remember and know judgments over learning proposed by Reder et al. (2000). Their representation of the relevant memory strength dimensions is quite similar to ours: They argue that one dimension of strength is the episodic strength of an item (akin to our specific-strength dimension) and that the other relevant dimension is semantic strength (similar to our global strength dimension). Full comparison of these alternate decision criteria within the two-dimensional framework requires different sets of experimental data. If the first judgment in a two-response task is "remember" versus "not remember," the process-pure model is the most natural one, whereas the sum–difference model makes more sense if the first response has "old" as an option. At least one other R-K experiment (Yonelinas & Jacoby, 1995, Experiment 3) follows the Reder et al. response pattern. In general, small changes in procedure can be expected to require changes in decision bounds within the two-dimensional model.

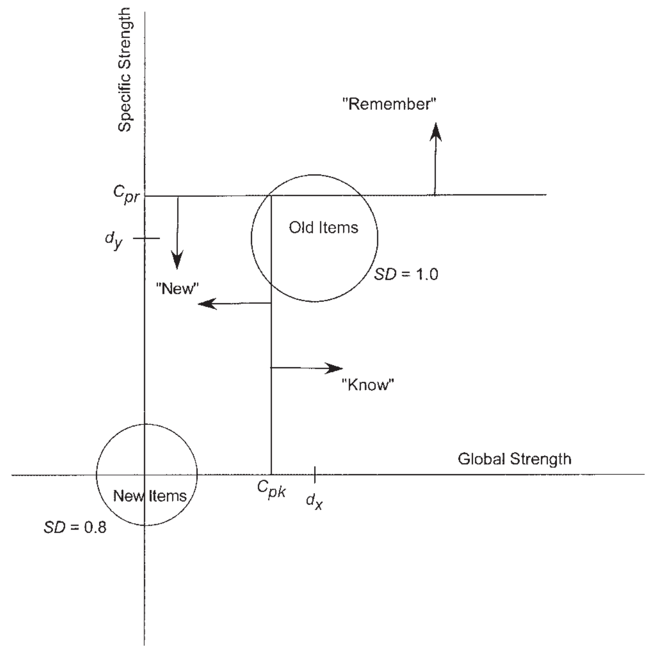

Figure 12. Two-dimensional signal-detection model of remember–know judgments with process-pure decision bounds. Old and New items differ in both specific and global strength, as in Figure 4. Observations of more than criterial specific strength lead to remember responses. Those at a lower level of specific strength are compared against a global strength criterion that divides know from new responses. Circles represent equal-likelihood contours from bivariate distributions. d_x = diagnosticity of global information; d_y = diagnosticity of specific information; C_{pr} = distance of the remember/not-remember decision bound from the mean of the New distribution; C_{pk} = distance of the know/not-know decision bound from the mean of the New distribution.

Versions that produce correlated remember responses and old–new confidence. Yonelinas (2001) reported that remembering was highly correlated with the confidence with which an item was called "old": Almost all remember judgments occurred with highest confidence. In Experiment 1, we also found that the probability of a remember response was correlated with confidence that the item is old (see Table 4). This aspect of the data is problematic for the simplest version of STREAK, in which the proportion of remember responses following old decisions is constant.

We evaluated several means of obtaining such a correlation in the two-dimensional model: (a) using nonorthogonal decision bounds, (b) relaxing the assumption of equal variance in the *x*- and *y*-directions, (c) assuming that items' global and specific memory strengths were correlated, and (d) allowing the (orthogonal, linear) R-K and old–new criteria, C_r and C_o, to move simultaneously. Simulation work revealed that only the last of these approaches has the potential to produce the desired correlation while continuing to describe the entire set of previously fit data (although combinations of the other three approaches may have some success). This more complex version of STREAK also appears able to account for the observation in Experiments 1A and 1B that the O-N rating/"know" data were slightly more linear than predicted by the simplest version of the model. Unfortunately, that version has too

Table 4
Probability of a Remember Response at Each Level of Confidence That the Item Was Old

	Experiment 1A		Experiment 1B	
Confidence	Target	Lure	Target	Lure
1	.94	.80	.94	.86
2	.73	.44	.81	.62
3	.58	.30	.65	.48
4			.57	.33
5			.47	.25

many parameters to be evaluated with the current data; more complex experiments are required.

The correlation of remember responses with old–new confidence is a challenge for the simplest version of the model, but it is not devastating because the correlation is not a pervasive result. A meta-analysis of six response bias experiments that influenced participants' willingness to say "old" and also collected R-K judgments has revealed a constant ratio of remember hit rate to old hit rate (Rotello et al., 2004). The constancy is exactly predicted: As the old–new criterion is adjusted to reflect different expectations about the number of Old items on the test list, for example, the rate of both old and remember responses to Old and New items changes by the same proportion. It is not clear what determines whether remember response rates are correlated with old–new confidence, but one factor that appears to play a role is whether the judgments are collected from the same or different participants. Within-participant manipulations of old–new bias, as in the experiments reported here, reveals a correlation; between-participants manipulations do not.

The one-dimensional model, in either the equal- or unequal-variance form, faces the opposite pattern of difficulty fitting the data: It can describe the correlation of remember and old responses by allowing both criteria to move when old–new bias changes, but the constant proportionality of the remember rates across old–new biases requires that the criteria move in a synchronized fashion. Rotello et al. (2004) have shown that neither version of the one-dimensional model can easily describe this aspect of the data.

Other Two-Dimensional Models of Memory

Multidimensional decision-theoretic models have a long history in cognition and perception (Tanner, 1956) and have been increasingly popular since the advent of general recognition theory (Ashby & Townsend, 1986). Extending SDT to multiple dimensions has been helpful in many research areas, such as categorization and speech perception, for which distinctions between sensitivity and bias matter but the stimulus sets are clearly not one-dimensional.

An important difference between STREAK and most other general recognition theory applications is that our interest is in just two stimulus classes, and thus there are only two distributions in the decision space. In the past, it has been the presence of a third class of stimuli (for example, two different stimulus lists in source memory experiments, e.g., Banks, 2000) that has necessitated a second dimension. In our case, the point of having two dimensions is to allow for the possibility that different memory judgments call on different aspects of the same multidimensional structure. Projection of distributions onto multiple decision axes in the space is a mathematical metaphor for this idea.

It is also important to point out that unlike other multidimensional representations of memory judgments, the two axes in our model do not represent particular stimulus dimensions (such as visual vs. auditory information). Such accounts may propose any number of dimensions, each representing an individual stimulus or event feature. In our two-dimensional model, the availability of any such features is represented by the specific-strength axis, but it is always contrasted to global strength, which reflects the overall availability of information in memory.

Behavioral Dissociations Between Remembering and Knowing

Much of the interest in the R-K paradigm arises from dissociable effects on remember and know judgments that have been found for experimental variables. In this section, we consider the characteristics and possible interpretations of such findings.

Dissociation refers to experimental manipulations that affect different aspects of performance in different ways. In the R-K paradigm, the influence of an experimental variable on the proportions of remember and know responses to Old items is typically examined. For example, depth-of-processing instructions increase remember responses but decrease know responses (Gardiner, 1988; Rajaram, 1993); changing stimulus materials can affect remember responses and leave know responses unchanged (see Rajaram, 1993, for pictures vs. words; see Gardiner & Java, 1990, for words vs. nonwords); and type of rehearsal can affect know responses but not remember responses (Gardiner, Gawlick, & Richardson-Klavehn, 1994). Several excellent reviews of such experiments have been published in recent years (e.g., Gardiner & Richardson-Klavehn, 2000; Rajaram, 1999; Rajaram & Roediger, 1997).

R-K dissociations are evaluated statistically by testing for interactions in an analysis of variance. Interactions in which both responses are affected in the same direction but to different degrees are usually not considered to be dissociations, but any other pattern is. Observed dissociations are then taken as evidence that the variable in question affects one and only one of the two processes (or two processes in opposite ways), a conclusion that the remember and know responses are process pure. The logic of this argument requires refinement. (See Dunn, 2004, for a discussion of R-K dissociations, and Dunn & Kirsner, 1988, for consideration of dissociation data more generally.)

A simple example is instructive. Consider the classic auditory detection experiment in which tones are sometimes but not always presented in noise and the listener responds "yes" (tone present) or "no." Hits (saying "yes" to the tone) and false alarms ("yes" for noise alone) are measured in two experimental conditions differing in the intensity of the tone. The plausible outcomes across the conditions are (a) hits increase and false alarms stay the same, (b) hits increase and false alarms decrease, (c) hits stay the same and false alarms decrease, and (d) hits and false alarms change in the same direction. By the logic of dissociation paradigms, outcomes a–c qualify as dissociations, but few would conclude that hits and false alarms measure distinct processes.

Rather than focusing on interactions, the usual interpretation of detection data examines measures of performance that are theoretically distinct, sensitivity and response bias. The well-known result (Green & Swets, 1966) is that variables that should affect sensitivity do so, and variables that should affect response bias but not sensitivity display that pattern. The logic differs from dissociation logic in two ways: First, the important finding is invariance rather than interaction; second, the appropriate measures of performance are theory-driven. Invariance plays some role in the conventional analysis because dissociation is inferred when one type of response is invariant while the other changes; the question is whether raw proportions of response types are informative about underlying processes. In auditory detection, a long history supports the conclusion that the hit and false-alarm rates are less instructive than sensitivity and bias estimates (which are the transformed difference and sum of these proportions; see Macmillan & Creelman, 1991, Chapter 1). But, this conclusion is derived from experiments in which experimental variables affect sensitivity and bias more predictably than hits and false alarms and from a theory that justifies sensitivity and bias parameters.

All models of the R-K paradigm convert the response pattern into sensitivity and bias parameters. The dual-process model asserts that the operation of recollective and familiarity processes can be independently measured by sensitivity parameters abstracted from ROC data. In rating experiments, therefore, variables that affect one or the other of these processes should be diagnosable from changes in aspects of ROC curves. Yonelinas (2001) has offered some support for this argument from his own rating data; the theory does not offer a strategy for teasing apart the two processes in nonrating experiments.

The one-dimensional model asserts that remember responses depend on the location of the remember criterion and know responses on the locations of both the remember and old criteria. Nonrating data can be used to evaluate this claim. A single sensitivity parameter reflects memory accuracy, and a change in sensitivity can also affect remember responses and/or know responses. We are not aware of any systematic attempts to study the influence of independent variables on the parameters of the one-dimensional model.

In the two-dimensional model, two sensitivity and two bias parameters combine to determine the response proportions. Therefore, STREAK allows us to ask not only whether the effects of variables used in R-K experiments are on sensitivity or bias but on what kind of sensitivity (specific or global) and bias (old–new or R-K).

The answers to these questions can be gleaned from Appendix A. Consider the examples mentioned at the beginning of this section. The depth-of-processing manipulation used by Gardiner (1988) and Rajaram (1993)—associates versus rhymes—affected both types of sensitivity. In Rajaram's data the R-K criterion is also affected. Contrasting pictures and words primarily affected d_x (Rajaram, 1993), whereas contrasting words and nonwords primarily affected participants' willingness to say "remember" (C_r; Gardiner & Java, 1991). Type of rehearsal had modest effects on d_x, d_y, and C_r (Gardiner et al., 1994).

These analyses are typical in that when STREAK is fit to data, more than one parameter changes between conditions. Explaining what appears to be a clean dissociation with a combination of effects may seem to be a step backward, but a quantitative model is needed to interpret the data. Whether changes in response rates arise from sensitivity or bias changes is of substantive importance; so, the exact pattern of hits, false alarms, remember hits, and remember false alarms is informative.

One important generalization is permitted by the parameter estimates in Appendix A: Most manipulations that have been studied mainly affect sensitivity (with or without also affecting bias). This conclusion is consistent with past research on the variables studied, which reveals a preponderance of accuracy effects. Before inferring qualitatively different processes, we must find measures that reflect them. The dissociation literature implicitly assumes that the appropriate statistics are the simple proportions of remember and know responses. The two-dimensional model claims that the theoretically important statistics depend on combinations of remember and know proportions. (Dunn, 2004, reached a similar conclusion based on the one-dimensional model.)

Neurological Evidence

Another way of approaching the issue of whether judgments of remembering and knowing tap separate memory systems or processes is to consider populations of individuals in which one of those putative systems is unavailable. Amnesic patients or older adults, who are assumed to have a reduced ability to use recollective processes or explicit memory systems, are ideal participants in studies of this sort. Similarly, the application of certain drugs is known to cause a temporary impairment of explicit memory in normal populations. Our analysis of these data parallels our analysis of the dissociation literature because the primary evidence is the same: Remember responses decrease while know responses either increase or are constant, a standard dissociation pattern.

A potentially more compelling type of evidence for distinct processes of remembering and knowing comes from the assessment of the brain activity associated with remember or know responses. Recent evoked response potential (ERP) studies have provided further support for the distinction between familiarity and recollection in recognition memory. In particular, familiarity-based responses have been associated with early activation in frontal recording sites (e.g., T. Curran, 1999, 2000), whereas recollective responses have been associated with later activation in parietal recording sites (see Allan et al., 1998, for a review). In addition, early anterior activation and later posterior activation have been observed in conjunction with know and remember responses, respectively (e.g., Duzel, Yonelinas, Mangun, Heinze, & Tulving, 1997; Rugg, Schloerscheidt, & Mark, 1998). The time course of these evoked potentials is consistent with the standard behavioral finding that familiarity influences memory earlier in a recognition judgment than recollection (e.g., T. Curran, 1999, 2000; Rotello & Heit, 2000). Moreover, the data indicate that the early activation pattern always occurs when a familiar test item is presented, including those that might later be rejected on the basis of recollection. In other words, the later activation appears to be supplementary to the earlier activation, just as recollection has been thought to supplement familiarity in information-processing terms.

A process-pure interpretation of these dissociations suggests that the assessment of familiarity (and thus the experience of knowing) is a fast-acting function of the frontal cortex but that recollection (and remembering) is a late-onset parietal function. From the perspective of the STREAK model, it is possible that the "signals" of global and specific memory strength are separately generated in these two regions. Although the model does not specify a time course for the contributions of global and specific memory strength to the decision process, it is possible that temporal dissociations of familiarity and recollection, and of activity in their associated brain regions, reflect the relatively late onset of the specific-strength signal. According to this hypothesis, speeded R-K judgments would be primarily influenced by global strength, but slower judgments (as in most behavioral R-K experiments) would be influenced by a joint assessment of global and specific strength, as indicated by the two-dimensional model.

The ERP dissociations are also consistent with the possibility that memory strength initially arises as one global signal, observed as the early frontal activation, but that parietal regions are subsequently recruited to deconstruct that signal into more specific components. In this account, familiarity and recollection reflect the availability of the same information in memory, but familiarity is a global view of that information whereas recollection highlights particular details.

An important feature of STREAK is the joint assessment of the two signals against an integrated criterion. This decision process may be associated with activation in the frontal cortex or elsewhere; because this activation would occur for both remember and know responses, however, it would not necessarily appear in subtractive imaging techniques such as standard ERP protocols. In short, the ERP data provide general support for the distinction between familiarity and recollection, or global and specific memory strength, but only tantalizing clues to their functional organization.

Conclusion

We evaluated three quantitative models of R-K judgments, two forms of the one-dimensional SDT model (Donaldson, 1996; Hirshman & Master, 1997) and a new two-dimensional SDT model, STREAK. In addition, we evaluated an extension of the dual-process model of recognition (Yonelinas, 1994) to R-K judgments. From our analyses, four main conclusions are clear. First, remember responses are not simply high-confidence old decisions, as is assumed in the one-dimensional model. A meta-analysis of 373 experimental conditions in the R-K literature demonstrated that the slope of the two-point *z*ROCs for these data is inconsistent with that observed in standard old–new recognition experiments.

Second, remember responses do not result from a high-threshold process, as is assumed by the dual-process model, although particular experimental procedures can lead to that conclusion (e.g., Rotello et al., 2004; Yonelinas, 2001; Yonelinas et al., 1996). This conclusion is supported by the finding in Experiment 2 that participants are able to rate their relative sense of remembering and knowing on a continuum.

Third, old–new and R-K judgments measure the sum and difference, respectively, of two different aspects of the memory strength of probes, global and specific strength. This conclusion is supported by the success of the two-dimensional model, which characterizes neither old–new nor R-K judgments as process pure. Application of the model to other tasks, such as those that ask participants to first decide whether they remember a probe and, failing that, to rate their degree of knowing or the probe's familiarity, may involve a different set of decision bounds.

Finally, and more generally, interpreting the empirical dissociations observed in remember and know judgments requires a specific quantitative model of the task. The finding that one type of response (remember) changes without necessarily affecting the alternative type of response (know) is not sufficient to conclude that the judgments tap different underlying memory processes or systems. In our model, both remember and know judgments depend on a combination of specific and global memory strength of the test probe. It is only the relative contributions of these two types of information that result in a decision that an item is remembered (if relatively more specific information can be retrieved) or known (if relatively more global memory strength is present). Determining whether STREAK, or an alternative model, provides the best explanation of any particular R-K task requires detailed quantitative analysis of the full set of responses.

References

References marked with an asterisk indicate studies included in the meta-analysis in Appendix A.

Allan, K., Wilding, E. L., & Rugg, M. D. (1998). Electrophysiological evidence for dissociable processes contributing to recollection. *Acta Psychologica, 98,* 231–252.

Ashby, F. G., & Townsend, J. T. (1986). Varieties of perceptual independence. *Psychological Review, 93,* 154–179.

Atkinson, R. C., & Juola, J. F. (1974). Search and decision processes in recognition memory. In D. H. Krantz, R. C. Atkinson, R. D. Luce, & P. Suppes (Eds.), *Contemporary development in mathematical psychology: Learning, memory, and thinking* (pp. 243–293). New York: Freeman.

Banks, W. P. (2000). Recognition and source memory as multivariate decision processes. *Psychological Science, 11,* 267–273.

*Blaxton, T. A., & Theodore, W. H. (1997). The role of the temporal lobes in recognizing visuospatial materials: Remembering versus knowing. *Brain and Cognition, 35,* 5–25.

Brainerd, C. J., Reyna, V. F., & Mojardin, A. H. (1999). Conjoint recognition. *Psychological Review, 106,* 160–179.

Clark, S. E., & Gronlund, S. D. (1996). Global matching models of memory: How the models match the data. *Psychonomic Bulletin & Review, 3,* 37–60.

*Clarys, D., Isingrini, M., & Gana, K. (2002). Mediators of age-related differences in recollective experience in recognition memory. *Acta Psychologica, 109,* 315–329.

Coltheart, M. (1981). The MRC Psycholinguistic Database. *Quarterly Journal of Experimental Psychology: Human Experimental Psychology, 33*(A), 497–505.

*Conway, M. A., & Dewhurst, S. A. (1995). Remembering, familiarity, and source monitoring. *Quarterly Journal of Experimental Psychology: Human Experimental Psychology, 48*(A), 125–140.

*Curran, H. V., Gardiner, J. M., Java, R. I., & Allen, D. (1993). Effects of lorazepam upon recollective experience in recognition memory. *Psychopharmacology, 110,* 374–378.

*Curran, H. V., & Hildebrandt, M. (1999). Dissociative effects of alcohol on recollective experience. *Consciousness and Cognition: An International Journal, 8,* 497–509.

Curran, T. (1999). The electrophysiology of incidental and intentional retrieval: ERP old-new effects in lexical decision and recognition memory. *Neuropsychologia, 37,* 771–785.

Curran, T. (2000). Brain potentials of recollection and familiarity. *Memory & Cognition, 28,* 923–938.

Curran, T., & Hintzman, D. L. (1995). Violations of the independence assumption in process dissociation. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 21,* 531–547.

*Dewhurst, S. A. (2001). Category repetition and false recognition: Effects of instance frequency and category size. *Journal of Memory and Language, 44,* 153–167.

*Dewhurst, S. A., & Anderson, S. J. (1999). Effects of exact and category repetition in true and false recognition memory. *Memory & Cognition, 27,* 665–673.

*Dewhurst, S. A., & Conway, M. A. (1994). Pictures, images, and recollective experience. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 20,* 1088–1099.

*Dewhurst, S. A., Hitch, G. J., & Barry, C. (1998). Separate effects of word frequency and age of acquisition in recognition and recall. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 24,* 284–298.

*Dobbins, I. G., Kroll, N. E. A., & Liu, Q. (1998). Confidence–accuracy inversions in scene recognition: A remember–know analysis. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 24,* 1306–1315.

Donaldson, W. (1996). The role of decision processes in remembering and knowing. *Memory & Cognition, 24,* 523–533.

Dosher, B. A. (1984). Discriminating preexperimental (semantic) information from learned (episodic) associations: A speed–accuracy study. *Cognitive Psychology, 16,* 519–555.

Dunn, J. C. (2004). Remember–know: A matter of confidence. *Psychological Review, 111,* 524–542.

Dunn, J. C., & Kirsner, K. (1988). Discovering functionally independent mental processes: The principle of reversed association. *Psychological Review, 95,* 91–101.

Duzel, E., Yonelinas, A. P., Mangun, G. R., Heinze, H.-J., & Tulving, E. (1997). Event-related brain potential correlates of two states of conscious awareness in memory. *Proceedings of the National Academy of Sciences, USA, 94,* 5973–5978.

*Eldridge, L. L., Sarfatti, S., & Knowlton, B. J. (2002). The effect of testing procedure on remember-know judgments. *Psychonomic Bulletin & Review, 9,* 139–145.

*Gardiner, J. M. (1988). Functional aspects of recollective experience. *Memory & Cognition, 16,* 309–313.

*Gardiner, J. M., Gawlick, B., & Richardson-Klavehn, A. (1994). Maintenance rehearsal affects knowing, not remembering; elaborative rehearsal affects remembering, not knowing. *Psychonomic Bulletin & Review, 1,* 107–110.

*Gardiner, J. M., & Gregg, V. H. (1997). Recognition memory with little or no remembering: Implications for a detection model. *Psychonomic Bulletin & Review, 4,* 474–479.

*Gardiner, J. M., & Java, R. I. (1990). Recollective experience in word and nonword recognition. *Memory & Cognition, 18,* 23–30.

*Gardiner, J. M., & Java, R. I. (1991). Forgetting in recognition memory with and without recollective experience. *Memory & Cognition, 19,* 617–623.

*Gardiner, J. M., Java, R. I., & Richardson-Klavehn, A. (1996). How level of processing really influences awareness in recognition memory. *Canadian Journal of Experimental Psychology, 50,* 114–122.

*Gardiner, J. M., Kaminska, Z., Dixon, M., & Java, R. I. (1996). Repetition of previously novel melodies sometimes increases both remember and know responses in recognition memory. *Psychonomic Bulletin & Review, 3,* 366–371.

*Gardiner, J. M., & Parkin, A. J. (1990). Attention and recollective experience in recognition memory. *Memory & Cognition, 18,* 579–583.

Gardiner, J. M., Ramponi, C., & Richardson-Klavehn, A. (2002). Recognition memory and decision processes: A meta-analysis of remember, know, and guess responses. *Memory, 10,* 83–98.

Gardiner, J. M., & Richardson-Klavehn, A. (2000). Remembering and knowing. In E. Tulving & F. I. M. Craik (Eds.), *The Oxford handbook of memory* (pp. 229–244). New York: Oxford University Press.

*Gardiner, J. M., Richardson-Klavehn, A., & Ramponi, C. (1997). On reporting recollective experiences and "direct access to memory systems." *Psychological Science, 8,* 391–394.

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

Glanzer, M., Kim, K., Hilford, A., & Adams, J. K. (1999). Slope of the receiver-operating characteristic in recognition memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 25,* 500–513.

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

*Gregg, V. H., & Gardiner, J. M. (1991). Components of conscious awareness in a long-term modality effect. *British Journal of Psychology, 82,* 153–162.

*Guttentag, R. E., & Carroll, D. (1997). Recollection-based recognition: Word frequency effects. *Journal of Memory and Language, 37,* 502–516.

*Hicks, J. L., & Marsh, R. L. (1999). Remember-know judgments can depend on how memory is tested. *Psychonomic Bulletin & Review, 6,* 117–122.

*Hicks, J. L., Marsh, R. L., & Ritschel, L. (2002). The role of recollection and partial information in source monitoring. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 28,* 503–508.

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

*Hirshman, E., & Henzler, A. (1998). The role of decision processes in conscious recollection. *Psychological Science, 9,* 61–65.

Hirshman, E., & Hostetter, M. (2000). Using ROC curves to test models of recognition memory: The relationship between presentation duration and slope. *Memory & Cognition, 28,* 161–166.

*Hirshman, E., & Lanning, K. (1999). Is there a special association between self judgements and conscious recollection? *Applied Cognitive Psychology, 13,* 29–42.

Hirshman, E., & Master, S. (1997). Modeling the conscious correlates of recognition memory: Reflections on the remember-know paradigm. *Memory & Cognition, 25,* 345–351.

*Hockley, W. E., & Consoli, A. (1999). Familiarity and recollection in item and associative recognition. *Memory & Cognition, 27,* 657–664.

*Holmes, J. B., Waters, H., & Rajaram, S. (1998). The phenomenology of false memories: Episodic context and confidence. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 24,* 1026–1040.

*Huron, C., Danion, J.-M., Rizzo, L., Killofer, V., & Damiens, A. (2003). Subjective qualities of memories associated with the picture superiority effect in schizophrenia. *Journal of Abnormal Psychology, 112,* 152–158.

*Inoue, C., & Bellezza, F. S. (1998). The detection model of recognition using know and remember judgments. *Memory & Cognition, 26,* 299–308.

Jacoby, L. L. (1991). A process dissociation framework: Separating automatic from intentional uses of memory. *Journal of Memory and Language, 30,* 513–541.

Johnson, M. K., Hashtroudi, S., & Lindsay, D. S. (1993). Source monitoring. *Psychological Bulletin, 114,* 3–28.

*Kinoshita, S. (1995). The word frequency effect in recognition memory versus repetition priming. *Memory & Cognition, 23,* 569–580.

*Knowlton, B. J., & Squire, L. R. (1995). Remembering and knowing: Two different expressions of declarative memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 21,* 699–710.

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

*Lampinen, J. M., Copeland, S. M., & Neuschatz, J. S. (2001). Recollections of things schematic: Room schemas revisited. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 27,* 1211–1222.

*LeCompte, D. C. (1995). Recollective experience in the revelation effect: Separating the contributions of recollection and familiarity. *Memory & Cognition, 23,* 324–334.

*Lindsay, D. S., & Kelley, C. M. (1996). Creating illusions of familiarity in a cued recall remember-know paradigm. *Journal of Memory and Language, 35,* 197–211.

*Long, D. L., & Prat, C. S. (2002). Memory for Star Trek: The role of prior knowledge in recognition revisited. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 28,* 1073–1082.

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

Macmillan, N. A., & Creelman, C. D. (1991). *Detection theory: A user's guide*. New York: Cambridge University Press.

Macmillan, N. A., Rotello, C. M., & Miller, J. O. (in press). The sampling distributions of Gaussian ROC statistics. *Perception & Psychophysics*.

Macmillan, N. A., Rotello, C. M., & Verde, M. F. (2004). *On the importance of models in interpreting remember–know experiments: Comments on Gardiner et al.'s (2002) meta-analysis.* Manuscript submitted for publication.

Mandler, G. (1991). Your face looks familiar but I can't remember your name: A review of dual process theory. In W. E. Hockley & S. Lewandowsky (Eds.), *Relating theory and data: Essays on human memory in honor of Bennet B. Murdock* (pp. 207–225). Hillsdale, NJ: Erlbaum.

*Mangels, J. A., Picton, T. W., & Craik, F. I. M. (2001). Attention and successful episodic encoding: An event-related potential study. *Cognitive Brain Research, 11,* 77–95.

*Mather, M., Henkel, L. A., & Johnson, M. K. (1997). Evaluating characteristics of false memories: Remember-know judgments and memory characteristics questionnaire compared. *Memory & Cognition, 25,* 826–837.

*Meiser, T., & Broeder, A. (2002). Memory for multidimensional source information. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 28,* 116–137.

Mitchell, K. J., & Johnson, M. K. (2000). Source monitoring: Attributing mental experiences. In E. Tulving & F. I. M. Craik (Eds.), *The Oxford handbook of memory* (pp. 179–195). New York: Oxford University Press.

*Mogg, K., Gardiner, J. M., Stavrou, A., & Golombok, S. (1992). Recollective experience and recognition memory for threat in clinical anxiety states. *Bulletin of the Psychonomic Society, 30,* 109–112.

*Moscovitch, D. A., & McAndrews, M. P. (2002). Material-specific deficits in "remembering" in patients with unilateral temporal lobe epilepsy and excisions. *Neuropsychologia, 40,* 1335–1342.

*Norman, K. A., & Schacter, D. L. (1997). False recognition in younger and older adults: Exploring the characteristics of illusory memories. *Memory & Cognition, 25,* 838–848.

*Ochsner, K. N. (2000). Are affective events richly recollected or simply familiar? The experience and process of recognizing feelings past. *Journal of Experimental Psychology: General, 129,* 242–261.

*Parkin, A. J., Gardiner, J. M., & Rosser, R. (1995). Functional aspects of recollective experience in face recognition. *Consciousness and Cognition, 4,* 387–398.

*Parkin, A. J., & Russo, R. (1993). On the origin of functional differences in recollective experience. *Memory, 1,* 231–237.

*Parkin, A. J., & Walter, B. M. (1992). Recollective experience, normal aging, and frontal dysfunction. *Psychology and Aging, 7,* 290–298.

*Perfect, T. J., Williams, R. B., & Anderton-Brown, C. (1995). Age differences in reported recollective experience are due to encoding effects, not response bias. *Memory, 3,* 169–186.

*Pesta, B. J., Murphy, M. D., & Sanders, R. E. (2001). Are emotionally charged lures immune to false memory? *Journal of Experimental Psychology: Learning, Memory, and Cognition, 27,* 328–338.

Raaijmakers, J. G., & Shiffrin, R. M. (1992). Models for recall and recognition. *Annual Review of Psychology, 43,* 205–234.

*Rajaram, S. (1993). Remembering and knowing: Two means of access to the personal past. *Memory & Cognition, 21,* 89–102.

Rajaram, S. (1999). Assessing the nature of retrieval experience: Advances and challenges. In B. H. Challis & B. M. Velichkovsky (Eds.), *Stratification in cognition and consciousness* (pp. 255–275). Philadelphia: John Benjamins.

*Rajaram, S., & Geraci, L. (2000). Conceptual fluency selectively influences knowing. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 26,* 1070–1074.

Rajaram, S., & Roediger, H. L., III. (1997). Remembering and knowing as states of consciousness during retrieval. In J. D. Cohen & J. W. Schooler (Eds.), *Scientific approaches to consciousness* (pp. 213–240). Hillsdale, NJ: Erlbaum.

Ratcliff, R., Sheu, C.-F., & Gronlund, S. D. (1992). Testing global memory models using ROC curves. *Psychological Review, 99,* 518–535.

Ratcliff, R., Van Zandt, T., & McKoon, G. (1995). Process dissociation, single-process theories, and recognition memory. *Journal of Experimental Psychology: General, 124,* 352–374.

*Reder, L. M., Nhouyvanisvong, A., Schunn, C. D., Ayers, M. S., Angstadt, P., & Hiraki, K. (2000). A mechanistic account of the mirror effect for word frequency: A computation model of remember–know judgments in a continuous recognition paradigm. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 26,* 294–320.

Rotello, C. M., & Heit, E. (2000). Associative recognition: A case of recall-to-reject processing. *Memory & Cognition, 28,* 907–922.

Rotello, C. M., Macmillan, N. A., Reeder, J. A., & Wong, M. (2004). *Two kinds of response bias in the remember-know paradigm.* Unpublished manuscript, University of Massachusetts at Amherst.

Rotello, C. M., Macmillan, N. A., & Van Tassel, G. (2000). Recall-to-reject in recognition: Evidence from ROC curves. *Journal of Memory and Language, 43,* 67–88.

Rugg, M. D., Schloerscheidt, A. M., & Mark, R. E. (1998). An electrophysiological comparison of two indices of recollection. *Journal of Memory and Language, 39,* 47–69.

*Schacter, D. L., Verfaellie, M., & Anes, M. D. (1997). Illusory memories in amnesic patients: Conceptual and perceptual false recognition. *Neuropsychology, 11,* 331–342.

*Schacter, D. L., Verfaellie, M., & Pradere, D. (1996). Neuropsychology of memory illusions—False recall and recognition in amnesic patients. *Journal of Memory and Language, 35,* 319–334.

*Smith, M. E. (1993). Neurophysiological manifestations of recollective experience during recognition memory judgments. *Journal of Cognitive Neuroscience, 5,* 1–13.

*Strack, F., & Forster, J. (1995). Reporting recollective experiences: Direct access to memory systems? *Psychological Science, 6,* 352–358.

Tanner, W. P., Jr. (1956). Theory of recognition. *Journal of the Acoustical Society of America, 28,* 882–888.

*Tulving, E. (1985). Memory and consciousness. *Canadian Psychology, 26,* 1–12.

*Verfaellie, M., Cook, S. P., & Keane, M. M. (2003). Absence of size congruency effects in amnesic patients' recognition: A failure of perceptually based recollection. *Neuropsychology, 17,* 108–114.

*Wagner, A. D., Gabrieli, J. D. E., & Verfaellie, M. (1997). Dissociations between familiarity processes in explicit recognition and implicit perceptual memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 23,* 305–323.

*Wallace, W. P., Malone, C. P., & Spoo, A. D. (2000). Implicit word activation during prerecognition processing: False recognition and remember-know judgments. *Psychonomic Bulletin & Review, 7,* 149–157.

*Wippich, W. (1992). Implicit and explicit memory without awareness. *Psychological Research, 54,* 212–224.

Yonelinas, A. P. (1994). Receiver-operating characteristics in recognition memory: Evidence for a dual-process model. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 20,* 1341–1354.

Yonelinas, A. P. (1997). Recognition memory ROCs for item and associative information: The contribution of recollection and familiarity. *Memory & Cognition, 25,* 747–763.

*Yonelinas, A. P. (2001). Consciousness, control, and confidence: The 3 Cs of recognition memory. *Journal of Experimental Psychology: General, 130,* 361–379.

*Yonelinas, A. P., Dobbins, I., Szymanski, M. D., Dhaliwal, H. S., & King, L. (1996). Signal-detection, threshold, and dual-process models of recognition memory: ROCs and conscious recollection. *Consciousness and Cognition, 5,* 418–441.

*Yonelinas, A. P., & Jacoby, L. L. (1995). The relation between remembering and knowing as bases for recognition: Effects of size congruency. *Journal of Memory and Language, 34,* 622–643.

Appendix A

Slopes of the Two-Point zROC Data for Experimental Conditions Reported by Donaldson (1996) and Published Since His Survey

Study and experiment	Level or condition	Slope of zROC	Best fitting parameter values from the STREAK model			
			C_o	d_x	d_y	C_r
Experimental conditions summarized by Donaldson (1996)						
Conway & Dewhurst (1995)						
1	Perform	0.704	1.07	0.9	2.7	0.4
	Watch	0.674	1.07	0.8	1.4	0.2
	Imagine	0.742	1.07	0.6	1.4	−0.1
2	Perform	0.805	0.94	0.8	1.5	0.5
	Watch	**1.102**	0.94	0.7	1.1	0
	Imagine	0.998	0.94	0.6	1.2	0.1
H. V. Curran, Gardiner, Java, & Allen (1993)	Predrug	0.911	0.98	1.1	2.1	0.7
	1 hr postdrug	0.893	0.98	0.7	1.4	0.4
	3 hr postdrug	0.994	0.98	0.9	1.4	0.4
	5 hr postdrug	**1.206**	0.98	1.0	1.7	0.3
	Preplacebo	0.771	0.98	1.0	1.7	0.7
	1 hr postplacebo	0.518	0.98	1.0	2.2	0.9
	3 hr postplacebo	0.828	0.98	1.1	2.1	0.8
	5 hr postplacebo	0.950	0.98	1.2	2.3	0.9
Dewhurst & Conway (1994)						
1	Pictures	0.758	0.94	1.4	2.5	0.9
	Words	0.823	0.94	0.7	1.1	0
2	Pictures	0.656	1.12	1.1	2.1	0.6
	Words	0.835	1.12	1.5	2.7	0.9
4	Words name	0.913	0.65	0.7	1.3	0.1
	Pictures name	1.003	1.12	1.2	1.9	0.6
	Words rate	—				
	Pictures rate	**1.237**	1.12	1.2	1.8	0.4
5	High	0.615	1.12	0.9	2.2	0.7
	Low	0.231	1.03	0.7	2.0	0.9
Gardiner (1988)						
1	Associates	**2.191**	1.64	1.7	2.2	0.8
	Rhyme	**2.353**	1.50	1.2	1.4	0.7
2	Generate 1 hr	**2.029**	1.12	1.3	1.7	0.9
	Read 1 hr	**1.495**	1.12	0.7	0.9	0.6
	Generate 1 week	0.681	0.49	0.4	1.0	0.5
	Read 1 week	0.812	0.49	0.2	0.6	0.2
Gardiner et al. (1994)	Learn immediate	0.620	0.90	0.8	1.5	0.6
	Forget immediate	0.746	0.90	0.4	0.9	0.1
	Learn delay	0.919	0.90	0.8	1.3	0.2
	Forget delay	1.019	0.90	0.6	1.0	−0.1
Gardiner & Java (1990)						
1	LF	0.820	0.98	0.7	1.3	0.6
	HF	0.745	0.86	0.4	1.1	0.5
2	Word	0.605	0.83	0.4	0.8	0.3
	Nonword	**1.081**	0.86	0.5	0.7	−0.2
Gardiner & Java (1991)						
1	10 min	—				
	1 hr	0.722	1.18	0.9	1.8	0.3
	1 day	**1.321**	1.07	0.7	1.1	0
	1 week	**1.204**	0.90	0.5	0.8	0.1
2	1 week	**1.120**	0.83	0.5	0.7	0
	4 weeks	0.755	0.98	0.4	0.9	−0.1
	6 weeks	1.007	0.86	0.2	0.5	0
Gardiner & Parkin (1990)	Undivided attention	0.769	1.32	1.0	2.2	0.6
	Divided 1	0.551	1.12	0.7	2.0	0.4
	Divided 2	0.563	1.03	0.5	3.0	0.1
Gregg & Gardiner (1991)	Spoken	0.704	0.65	0.9	2.2	0.5
	Silent	0.662	0.65	0.5	3.0	0
	Spoken	0.815	0.62	0.8	1.8	0.3
	Silent	0.799	0.62	0.5	1.2	−0.1

(*Appendixes continue*)

Appendix A (*continued*)

Study and experiment	Level or condition	Slope of *z*ROC	Best fitting parameter values from the STREAK model			
			C_o	d_x	d_y	C_r
Experimental conditions summarized by Donaldson (1996) (*continued*)						
LeCompte (1995)						
2	Normal	0.973	0.33	0.5	0.9	0.4
	Revealed	0.953	0.12	0.4	0.7	0.1
Mogg, Gardiner, Stavrou, & Golombok (1992)	Threat	**1.275**	1.03	0.7	1.0	0
	Nonthreat	0.781	1.18	0.8	2.0	0.2
	Threat	**1.251**	0.94	0.7	0.9	−0.2
	Nonthreat	**1.560**	1.32	1.0	1.3	0.2
Parkin & Russo (1993)	Massed	0.875	0.59	0.8	1.4	0.4
	Spaced	0.765	0.59	0.9	1.9	0.8
Parkin & Walter (1992)						
1	Young	1.011	1.32	1.2	1.9	0.5
	Old	**1.200**	1.03	0.8	1.3	−0.5
2	Young	**2.030**	1.18	1.3	1.6	−0.1
	Middle	**1.887**	1.07	0.9	1.2	−0.5
	Old	**2.265**	1.07	1.0	1.1	−0.6
Rajaram (1993)						
1	Associate	0.630	0.80	1.0	2.3	0.8
	Rhyme	0.730	0.80	0.6	1.4	0
	Visual	0.631	0.80	0.8	1.7	0.4
	Auditory	0.631	0.80	0.8	1.7	0.4
2	Pictures	0.410	1.07	1.5	2.8	1.1
	Words	0.478	1.07	0.8	3.0	0.7
3	Repeated prime	0.680	0.59	0.6	1.2	0.3
	Unrelated prime	0.624	0.73	0.5	1.4	0.6
Smith (1993)		**1.309**	1.07	1.4	2.3	0.2
Tulving (1985)						
2	1 day	**1.169**	0.76	1.1	1.8	0.9
	1 week	**1.170**	0.49	0.5	0.7	0.1
Wippich (1992)						
1	Read	0.746	0.86	0.4	0.6	−0.3
	Generate	0.829	0.86	0.8	1.5	0.2
Experimental conditions not included in Donaldson (1996)						
Blaxton & Theodore (1997)						
1	Normal	**1.954**	0.37	0.8	0.7	−0.2
	Left hemisphere	0.823	0.10	0.3	0.8	0.4
	Right hemisphere	0.507	0.31	0.3	1.2	0.2
1A	Left, presurgery	0.682	0.10	0.3	0.9	0.4
	Left, postsurgery	**1.163**	0.24	0.4	0.7	0.2
	Right, presurgery	0.468	0.24	0.3	1.2	0.3
	Right, postsurgery	0.418	0.20	0.2	1.2	0.5
2	Normal, label	0.972	0.65	0.7	1.3	0.8
	Normal, line ctrl	**3.013**	0.49	0.9	0.2	−0.6
	Left, label	**1.992**	0.59	0.9	0.7	−0.4
	Left, line ctrl	**2.952**	0.65	1.0	0.5	−0.8
	Right, label	**1.539**	0.49	0.6	0.7	0.3
	Right, line ctrl	0.832	0.59	0.2	0.3	0
Clarys, Isingrini, & Gana (2002)	Young	1.050	1.10	0.7	0.9	0
	Old	**1.558**	1.10	0.7	0.8	−0.2
	Very old	**1.371**	0.80	0.5	0.6	−0.5
H. V. Curran & Hildebrandt (1999)	Alcohol, generate	0.803	1.18	0.7	1.5	0.1
	Alcohol, read	0.810	1.18	0.6	1.5	0
	Placebo, generate	0.714	0.80	0.9	1.8	0.4
	Placebo, read	0.728	0.80	0.6	1.7	0.1
Dewhurst (2001)						
1	1×	0.576	0.90	0.7	1.4	0.4
	4×	0.717	0.90	0.7	1.5	0.3
	8×	0.867	0.90	0.8	1.6	0.1
2	HF	0.758	1.20	0.9	1.6	0.3
	LF	0.606	1.20	1.0	1.7	0.6
Dewhurst & Anderson (1999)						
1	Spaced, 1×	—				
	Spaced, 4×	0.580	0.98	1.3	2.6	0.9

Appendix A (*continued*)

Study and experiment	Level or condition	Slope of zROC	Best fitting parameter values from the STREAK model C_o	d_x	d_y	C_r
Experimental conditions not included in Donaldson (1996) (*continued*)						
	Spaced, 8×	0.457	0.73	1.1	2.6	0.9
	Massed, 1×	—				
	Massed, 4×	0.634	0.94	1.0	2.3	0.9
	Massed, 8×	0.810	0.80	1.1	1.9	0.9
2	Spaced, 1×	—				
	Spaced, 4×	0.875	0.62	0.5	1.1	0
	Spaced, 8×	0.815	0.35	0.4	0.9	−0.1
	Massed, 1×	0.902	0.86	0.6	1.0	0.3
	Massed, 4×	0.608	0.37	0.3	1.1	0.3
	Massed, 8×	0.830	0.22	0.3	0.8	0
Dewhurst, Hitch, & Barry (1998)						
1	HF, early	—				
	HF, late	0.892	0.86	0.9	1.7	0.4
	LF, early	0.524	0.94	0.8	2.0	0.5
	LF, late	1.045	1.12	1.2	2.4	0.7
2	Early	0.999	0.86	0.7	1.2	0
	Late	0.745	0.65	0.7	1.6	0.3
Dobbins, Kroll, & Liu (1998)	Same context	0.885	0.42	0.5	1.1	−0.1
	Changed context	**1.075**	0.33	0.4	0.7	0.2
	Novel context	0.976	0.70	0.9	1.6	0.2
Eldridge, Sarfatti, & Knowlton (2002)						
1	One-step instructions	0.854	0.80	1.0	1.8	0.7
	Two-step instructions	1.347	1.20	1.1	1.4	0.3
2	One-step instructions	0.669	0.40	0.7	1.7	0.4
	Two-step instructions	0.862	1.00	0.9	1.8	0.3
Gardiner & Gregg (1997)						
1		**1.406**	0.62	0.6	0.7	−0.8
2		**1.959**	0.65	0.9	0.9	−0.8
3		**1.498**	0.62	0.7	0.7	−0.8
4		**1.369**	0.37	0.4	0.4	−1.2
Gardiner, Java, & Richardson-Klavehn (1996)						
1	Deep	0.674	0.24	1.0	2.4	0.7
	Shallow	0.802	0.24	0.3	0.5	−0.7
2	Deep	**1.081**	0.59	0.9	1.7	0
	Shallow	**1.446**	0.59	0.7	0.9	−0.8
3	Generate	0.953	0.62	1.0	2.1	−0.2
	Read	0.996	0.62	0.6	1.2	−0.8
Gardiner, Kaminska, Dixon, & Java (1996)						
1	1 trial	**1.193**	0.40	0.4	0.5	0
	2 trials	**1.093**	0.67	0.7	1.0	−0.2
	4 trials	**1.592**	0.90	1.2	1.6	−0.3
2	1 trial	**1.887**	0.76	0.8	0.6	−0.2
	2 trials	**1.383**	0.80	0.9	1.1	−0.3
	4 trials	**1.133**	1.12	1.1	2.1	0.1
3	Polish, 1 trial	0.819	0.02	0.4	0.8	−0.5
	Polish, 3 trials	0.935	0.02	0.7	1.4	−0.2
	Classical, 1 trial	0.951	0.10	0.5	0.8	−0.6
	Classical, 3 trials	**1.081**	0.20	0.9	1.5	0
Gardiner, Richardson-Klavehn, & Ramponi (1997)	HF, 50%	0.651	0.42	0.4	1.2	−0.1
	HF, 30%	0.662	0.59	0.4	1.4	−0.1
	LF, 50%	0.656	0.62	0.7	1.5	0.1
	LF, 30%	0.646	0.51	0.5	1.5	0.1
Guttentag & Carroll (1997)						
1	HF	0.683	0.59	0.5	1.1	0.1
	LF	0.769	0.76	0.8	1.5	0.4
Hicks & Marsh (1999)	O/N + R/K	0.931	0.65	0.7	1.2	0.2
	R/K/N	0.845	0.40	0.6	1.3	0.5
	O/N + R/K	0.946	0.94	1.0	1.8	0.3
	R/K/N	**1.254**	0.67	1.0	1.5	0.4
Hicks, Marsh, & Ritschel (2002)						
1	Heard	**1.244**	0.80	0.8	1.4	0.1
	Seen	**1.236**	0.80	0.8	1.3	0

(*Appendixes continue*)

Appendix A (*continued*)

Study and experiment	Level or condition	Slope of *z*ROC	Best fitting parameter values from the STREAK model C_o	d_x	d_y	C_r
Experimental conditions not included in Donaldson (1996) (*continued*)						
2	Generated	**1.126**	0.70	1.1	1.8	0.3
	Seen	0.999	0.70	0.5	0.9	−0.2
Hirshman & Henzler (1998)	2 s, 30%	**1.187**	0.73	0.6	0.9	−0.1
	2 s, 70%	0.999	0.08	0.5	0.8	−0.1
	0.5 s, 30%	**1.190**	0.73	0.3	0.2	−0.5
	0.5 s, 70%	1.034	0.08	0.3	0.3	−0.5
Hirshman & Lanning (1999)						
1	Self-relevance	**2.026**	0.98	1.6	2.3	0.7
	Trait relevant	**1.807**	0.98	1.3	2.0	0.5
2	Self-relevance	**1.374**	1.07	1.5	2.5	0.6
	Average American	**1.219**	1.07	1.2	2.0	0.4
3	Self-relevance	**1.115**	1.07	1.3	2.2	0.8
	Average American	**1.616**	1.07	1.8	2.7	0.9
4	Self-relevance	**1.453**	1.07	1.5	2.4	0.6
	Average American	**1.767**	1.07	1.6	2.4	0.5
Hockley & Consoli (1999)						
1	Pairs: Immediate	0.419	0.57	0.9	2.1	0.9
	Pairs: 30 min	0.481	0.51	0.8	1.9	0.9
	Pairs: 24 hr	0.549	0.31	0.3	1.2	0.5
	Words: Immediate	0.715	0.65	0.7	1.6	0.6
	Words: 30 min	0.746	0.44	0.6	1.3	0.6
	Words: 24 hr	0.842	0.31	0.2	0.6	0.1
2	Pairs: Immediate	0.442	0.54	0.8	1.9	0.8
	Pairs: 48 hr	0.626	0.33	0.3	1.0	0.2
	Pairs: 1 week	0.743	0.37	0.2	0.4	−0.2
	Words: Immediate	0.839	0.70	0.8	1.4	0.5
	Words: 48 hr	0.978	0.31	0.2	0.4	−0.2
	Words: 1 week	0.925	0.18	0.1	0.2	−0.2
Holmes, Waters, & Rajaram (1998)						
1	4 idea units	**1.193**	0.18	0.1	0.1	0.3
	3 idea units	0.928	0.29	0	0.2	0.4
	2 idea units	0.938	0.51	0	0.2	0.1
	1 idea unit	0.615	0.94	0.1	0.6	0.2
2	4 idea units	0.795	0.08	0	0.1	0.5
	3 idea units	1.069	0.29	0.1	0.1	0.3
	2 idea units	1.065	0.51	0.1	0.1	0.1
	1 idea unit	**1.341**	0.90	0.2	0.1	0
3	4 idea units	**1.162**	0.33	0.1	0.1	−0.1
	3 idea units	1.068	0.29	0.1	0	−0.1
	2 idea units	**1.070**	0.47	0.1	0.1	−0.2
	1 idea unit	**1.203**	0.86	0.2	0.3	−0.2
Huron, Danion, Rizzo, Killofer, & Damiens (2003)	Schizophrenic, picture	**1.471**	1.50	1.4	2.0	0.6
	Schizophrenic, words	1.049	1.50	0.5	1.0	0
	Normal, picture	0.686	1.20	1.4	2.4	0.9
	Normal, word	0.824	1.20	0.5	1.2	−0.2
Inoue & Bellezza (1998)						
1	1 week	**1.163**	0.18	0.4	0.6	−0.1
2	24 hr: Intact pair	0.837	0.31	0.5	1.0	0.3
	24 hr: Rearranged pair	1.050	0.31	0.4	0.6	−0.2
	24 hr: Old–new pair	1.022	0.33	0.4	0.6	−0.2
Kinoshita (1995)						
1	HF read	0.499	1.24	0.7	1.4	0.5
	LF read	0.533	1.24	1.0	2.5	0.8
	HF judge	0.617	1.24	0	0.2	−0.6
	LF judge	0.746	1.24	0	0.1	−1.2
2	HF read	**1.470**	0.67	0.6	0.6	0
	LF read	**2.076**	0.67	1.1	1.4	0.2
	HF judge	**2.983**	0.83	0.9	0.3	−0.8
	LF judge	**2.215**	0.83	0.7	0.2	−0.6
Knowlton & Squire (1995)						
1	10 min	0.977	1.23	1.0	1.5	0.5
	1 week	1.063	0.49	0.3	0.4	0
	Amnesic	**1.149**	0.72	0.2	0.2	0

Appendix A (*continued*)

Study and experiment	Level or condition	Slope of *z*ROC	Best fitting parameter values from the STREAK model C_o	d_x	d_y	C_r
Experimental conditions not included in Donaldson (1996) (*continued*)						
2	10 min	0.609	1.49	1.0	2.0	0.7
	1 week	1.005	0.97	0.4	0.8	0.2
3	10 min	**1.203**	1.20	0.8	1.0	0.4
	1 week	0.827	1.10	0.4	1.2	0
Lampinen, Copeland, & Neuschatz (2001)						
1	Intentional, typical	0.838	0	0.2	0.4	−0.3
	Intentional, atypical	0.252	1.30	0.9	1.8	1.4
	Incidental, typical	0.791	0	0	0.1	−0.7
	Incidental, atypical	0.177	1.20	0.5	2.4	1.7
2	Intentional, typical, immediate	0.787	0.50	0.5	0.9	0.7
	Incidental, typical, immediate	0.453	0.40	0.2	1.0	0.1
	Intentional, typical, delay	0.750	0	0.1	0.4	0
	Intentional, atypical, delay	0.382	1.20	0.8	2.0	1.3
	Incidental, typical, delay	1.050	−0.10	0.1	0.2	0.1
	Incidental, atypical, delay	0.487	1.20	0.3	1.2	1.4
Lindsay & Kelley (1996)						
1	Easy	0.821	−0.12	0.4	0.8	−0.1
	Hard	0.703	−0.02	0.1	0.6	−0.1
2	Shallow, easy	0.914	−0.22	0.4	0.7	0.1
	Shallow, hard	1.004	−0.04	0.3	0.5	−0.1
	Deep, easy	0.798	0.18	1.4	2.4	0.9
	Deep, hard	0.809	0.31	0.7	1.5	0.9
3	Uninformed, easy	0.855	−0.27	0.4	0.9	0.3
	Uninformed, hard	0.771	−0.02	0.4	0.8	0.3
	Informed, easy	0.740	−0.20	0.4	1.0	0.3
	Informed, hard	1.034	0.02	0.5	0.8	0.2
Long & Prat (2002)						
1	Expert, ST material	0.538	0.80	0.7	1.5	1.1
	Expert, psych	0.755	0.50	0.6	1.3	0.4
	Novice, ST material	0.962	0.60	0.6	1.1	0.6
	Novice, psych	0.898	0.40	0.6	1.0	0.3
Mangels, Picton, & Craik (2001)	Full attention	0.690	0.90	1.2	2.2	0.6
	Easy divided	0.559	0.90	0.9	2.0	0.5
	Hard divided	0.764	0.60	0.4	0.6	−0.5
Mather, Henkel, & Johnson (1997)	Blocked, one	**1.415**	0.67	0.8	1.3	0.4
	Blocked, two	0.887	0.83	0.9	1.5	0.1
	Random, one	0.396	0.35	0.4	1.8	0.5
	Random, two	0.826	0.35	0.6	1.1	0.4
Meiser & Broeder (2002)						
2	Upper position, small font	**1.224**	1.00	0.9	1.2	0.3
	Upper position, large font	**1.392**	1.00	0.9	1.4	0.3
	Lower, small	**1.180**	1.00	0.9	1.2	0.4
	Lower, large	**1.304**	1.00	0.9	1.5	0.4
Moscovitch & McAndrews (2002)	Ctrl, words, perceptual	1.031	0.70	0.7	1.2	0.3
	Ctrl, words, conceptual	0.997	0.70	1.0	1.6	0.6
	Ctrl, faces, perceptual	**1.175**	0.90	0.6	0.7	−0.4
	Ctrl, faces, conceptual	**1.081**	0.90	1.0	1.8	0.4
	LT, words, perceptual	0.886	0.30	0.3	0.5	0.2
	LT, words, conceptual	1.007	0.30	0.4	0.7	0.3
	LT, faces, perceptual	1.013	0.70	0.4	0.5	0
	LT, faces, conceptual	**1.131**	0.70	0.9	1.6	0.6
	LT preop, words, perceptual	0.972	0.10	0.3	0.5	0.3
	LT preop, words, conceptual	1.026	0.10	0.3	0.6	0.4
	LT preop, faces, perceptual	0.674	0.70	0.3	0.6	0.3
	LT preop, faces, conceptual	0.908	0.70	0.8	1.5	0.8
	LT postop, words, perceptual	0.826	0.40	0.2	0.6	0.2
	LT postop, words, conceptual	0.984	0.40	0.5	0.8	0.2
	LT postop, faces, perceptual	**1.349**	0.70	0.5	0.5	−0.2
	LT postop, faces, conceptual	**1.469**	0.70	1.1	1.6	0.4
	RT, words, perceptual	1.007	0.50	0.3	0.4	0.3
	RT, words, conceptual	**1.220**	0.50	0.7	1.2	0.7
	RT, faces, perceptual	0.891	0.50	0.4	0.7	0.2
	RT, faces, conceptual	**1.134**	0.50	0.7	1.0	0.2
	RT preop2, words, perceptual	0.942	0.40	0.2	0.4	0.4
	RT preop2, words, conceptual	**1.132**	0.40	0.5	0.8	0.6

(*Appendixes continue*)

Appendix A (*continued*)

Study and experiment	Level or condition	Slope of *z*ROC	Best fitting parameter values from the STREAK model C_o	d_x	d_y	C_r
Experimental conditions not included in Donaldson (1996) (*continued*)						
	RT preop2, faces, perceptual	0.606	0.50	0.4	1.1	0.6
	RT preop2, faces, conceptual	0.572	0.50	0.6	1.5	0.9
	RT postop2, words, perceptual	0.878	0.60	0.3	0.7	0.3
	RT postop2, words, conceptual	**1.095**	0.60	0.9	1.6	0.7
	RT postop2, faces, perceptual	**1.264**	0.50	0.3	0.4	−0.1
	RT postop2, faces, conceptual	**1.556**	0.50	0.7	0.8	−0.1
Norman & Schacter (1997)						
1	Normal, explain	0.830	1.18	1.1	2.2	0.5
	Normal, no explanation	0.794	1.12	1.1	2.1	0.4
	Elderly, explain	**2.022**	0.80	1.0	1.2	0.6
	Elderly, no explanation	0.966	0.83	0.8	1.5	0.6
Ochsner (2000)						
1	Negative photos	1.036	0.94	1.0	1.8	0.4
	Neutral photos	**1.106**	0.98	0.9	1.6	0
	Positive photos	1.061	0.80	0.9	1.5	0.1
	High arousal	0.966	0.80	0.9	1.7	0.3
	Medium arousal	1.065	1.03	1.0	2.0	0.3
	Low arousal	1.015	0.90	0.9	1.3	0
2	Negative photos	**1.150**	0.76	0.8	1.4	0.3
	Neutral photos	**1.227**	0.94	0.7	1.1	−0.1
	Positive photos	**1.467**	0.76	0.8	1.0	0
	High arousal	**1.111**	0.65	0.8	1.2	0.3
	Medium arousal	**1.563**	1.03	1.0	1.2	0
	Low arousal	**1.120**	0.86	0.7	1.0	−0.2
3	Negative photos	0.918	0.76	0.9	1.7	0.4
	Neutral photos	1.011	0.76	0.7	1.4	0
	Positive photos	0.744	0.67	0.7	2.1	0.1
	High arousal	0.772	0.67	0.9	1.7	0.4
	Medium arousal	**1.223**	0.86	0.9	1.4	0.2
	Low arousal	0.877	0.76	0.7	1.6	−0.1
Parkin, Gardiner, & Rosser (1995)						
1	Full attention	**1.160**	0.86	1.0	1.7	0.4
	Divided	**1.134**	0.62	0.6	1.0	0.2
2	Massed	**1.083**	0.62	0.8	1.2	−0.1
	Spaced	**1.098**	0.62	1.0	1.8	0.3
Perfect, Williams, & Anderton-Brown (1995)						
2	Deep	**1.405**	1.50	1.5	1.9	0.9
	Shallow	**1.078**	1.12	0.7	1.3	0.3
Pesta, Murphy, & Sanders (2001)						
1	Related, unemotional	0.840	−0.30	0.3	0.8	0.7
	Related, emotional	**5.741**	0.70	1.3	1.3	0.7
	Unrelated, unemotional	**1.361**	0.40	0.5	0	−1.0
	Unrelated, emotional	**2.224**	1.30	0.7	0.5	−0.6
2	Related, unemotional	0.848	−0.60	0.2	0.4	0.8
	Related, emotional	**1.240**	0.20	0.6	1.0	1.2
	Unrelated, unemotional	0.808	0.40	0.4	0	−0.6
	Unrelated, emotional	**1.553**	0.90	0.3	0.1	−0.2
3	Related, unemotional	0.647	−0.60	0.1	0.5	0.5
	Related, emotional	**1.574**	0.10	0.6	0.8	0.6
	Unrelated, unemotional	0.860	0.20	0.2	0	−0.8
	Unrelated, emotional	0.778	0.70	0.1	0.5	0
4	Related, unemotional	0.930	−0.50	0.1	0.3	0.7
	Related, emotional	**2.012**	0.40	0.9	1.2	0.9
	Unrelated, unemotional	0.755	0.30	0.1	0	−1.0
	Unrelated, emotional	**1.655**	1.10	0.4	0.4	−0.5
Rajaram & Geraci (2000)	Related prime	0.994	0.80	0.8	1.2	0.2
	Unrelated prime	1.028	1.03	0.9	1.3	0.3
Reder et al. (2000)						
1	HF	**1.097**	0.80	1.4	2.4	0.7
	LF	**1.524**	1.24	2.0	3.0	0.9
2	HF	**1.245**	0.86	1.7	3.0	0.9
	LF	0.679	1.24	2.0	3.3	1.3
3	HF	0.765	0.62	0.3	0.6	−0.5
	LF	0.765	0.83	0.6	1.3	0

Appendix A (*continued*)

Study and experiment	Level or condition	Slope of *z*ROC	Best fitting parameter values from the STREAK model: C_o	d_x	d_y	C_r
Experimental conditions not included in Donaldson (1996) (*continued*)						
Schacter, Verfaellie, & Anes (1997)						
2	Ctrl: High associates	—				
	Ctrl: Perceptual	—				
	Amnesic: High associates	**1.720**	0.62	0.5	0.4	−0.2
	Amnesic: Perceptual	0.928	0.86	0.5	0.9	−0.3
Schacter, Verfaellie, & Pradere (1996)						
1	Amn study + recall	0.936	0.33	0.2	0.5	0.2
	Amn study + arith	0.860	0.33	0.1	0.4	0.2
	Amn critical lure study + recall	0.718	0.14	0.2	0.7	0.4
	Amn critical lure study + arith	0.901	0.14	0.2	0.3	0
	Ctrl study + recall	0.755	0.73	1.0	2.0	0.9
	Ctrl study + arith	0.627	0.73	1.0	1.9	0.9
	Ctrl critical lure study + recall	0.638	0.44	0.8	1.7	0.9
	Ctrl critical lure study + arith	0.405	0.44	1.1	2.2	1.1
Strack & Forster (1995)						
1	HF, 50%	0.646	0.51	0.4	1.5	−0.1
	HF, 30%	0.936	1.06	0.6	0.9	−0.1
	LF, 50%	0.670	0.53	0.7	1.6	0.2
	LF, 30%	0.871	0.82	0.7	1.3	0.2
2	Short test	**1.413**	1.24	1.6	2.6	0.5
	Long test	1.068	0.44	0.6	1.0	−0.3
Verfaellie, Cook & Keane (2003)						
1	Same size	**1.537**	0.40	0.6	0.8	0.4
	Different size	**1.496**	0.40	0.4	0.4	0.1
Wagner, Gabrieli, & Verfaellie (1997)						
4	Pictures	0.787	0.76	1.5	2.7	0.9
	Words	0.775	0.76	0.5	0.9	−0.1
	Heard	0.683	0.76	0.7	1.8	0.5
Wallace, Malone, & Spoo (2000)						
2	Words	**1.108**	0.98	1.1	1.7	0.2
	Early change	**1.270**	0.29	0.7	1.0	0.2
	Late change	**1.237**	0.20	0.6	0.9	0.2
	Words	**1.150**	1.03	0.9	1.7	0.1
	Early change	0.959	0.76	0.8	1.3	0.1
	Late change	0.908	0.62	0.7	1.2	0.1
Yonelinas (2001)						
1	Full attention	0.734	0.50	0.8	1.7	0.1
	Divided	0.814	0.50	0.6	1.4	−0.3
2a	Full attention	0.603	0.80	1.0	2.1	0.5
	Divided	0.646	0.80	0.8	1.5	0.1
2b	Full attention	0.728	0.50	0.4	1.2	−0.3
	Divided	0.825	0.50	0.4	0.6	−0.7
2c	Full attention	0.696	0.50	0.6	1.2	0
	Divided	0.795	0.50	0.4	0.8	−0.4
3	Deep	0.724	0.60	0.8	1.9	0.5
	Shallow	0.840	0.60	0.4	0.7	−0.4
Yonelinas et al. (1996)						
2	Twice	0.643	0.35	0.6	1.3	0.2
	Once	0.664	0.35	0.4	1.0	0
Yonelinas & Jacoby (1995)						
2	Size congruent	0.800	0.12	0.6	1.2	0.1
	Size incongruent	0.859	0.12	0.4	0.8	−0.2
3	Size congruent	0.689	0.22	0.5	1.4	0
	Size incongruent	0.657	0.22	0.4	0.9	−0.3

Note. Slopes greater than the maximum slope observed in the old–new recognition literature (i.e., 1.07) are in bold. Dashes indicate conditions for which the slope of the normal-normal receiver operating characteristic (*z*ROC) could not be calculated. STREAK = sum–difference theory of remembering and knowing; C_o = distance of the old–new decision bound from the mean of the New distribution; d_x = mean of the Old distribution on the global strength axis; d_y = mean of the Old distribution on the specific strength axis; C_r = distance of the remember–know decision bound from the mean of the Old distribution; LF = low frequency, rare word; HF = high frequency, common word; Massed = massed repetitions; Spaced = distributed repetitions; ctrl = control; 1× = repeated once; 4× = repeated four times; 8× = repeated eight times; O/N + R/K = sequential old–new and remember–know testing; R/K/N = trinary decision remember–know–new; 30% = told 30% of test items were Old; 50% = told 50% of test items were Old; 70% = told 70% of test items were Old; ST material = Star Trek materials; psych = psychology materials; LT = left temporal; RT = right temporal; Amn = amnesic; arith = arithmetic. For details on the experimental conditions, please see the original articles.

(*Appendixes continue*)

Appendix B

Equations Relating Response Probabilities to STREAK Parameters

Figure 4 displays the effects of Old and New items in a space of two dimensions, global memory strength (x-axis) and specific memory strength (y-axis). Each stimulus class produces a bivariate normal distribution: New items lead to mean strengths of (0, 0) and standard deviation of s on both dimensions, whereas Old items have means (d_x, d_y) and a standard deviation of 1 on each dimension. There are two decision bounds, both linear. To choose between old and new, the respondent uses a line with slope $-d_y/d_x$; if that response is old, then remember and know responses are determined by an orthogonal bound with slope d_x/d_y.

It is useful to project the bivariate distributions onto decision axes perpendicular to the decision bounds. Let the orthogonal distance from the mean of the New distribution to the old–new bound be C_o. Then, because the standard deviation is s, the false-alarm rate is the proportion of the projected New distribution that falls above the old–new bound and equals

$$P(\text{“old”}|\text{New}) = \Phi\left(\frac{-C_o}{s}\right). \tag{B1}$$

The hit rate is the proportion of the Old distribution, along the same decision axis, that falls above the old–new bound and equals

$$P(\text{“old”}|\text{Old}) = \Phi\left(\frac{2d_x d_y}{\sqrt{d_x^2 + d_y^2}} - C_o\right). \tag{B2}$$

The remember false-alarm rate is the proportion of the volume under the New distribution that is above both the old–new and the R-K decision bounds. Because these two bounds are orthogonal, this proportion can be found by multiplying the false-alarm rate (Equation B1) by the proportion of the New distribution above the R-K bound. Calling the orthogonal distance from the Old mean to this decision bound C_r, the orthogonal distance is

$$\frac{1}{s}\left(\frac{d_x^2 - d_y^2}{\sqrt{d_x^2 + d_y^2}} + C_r\right).$$

This fraction of the volume, multiplied by the false-alarm rate, is therefore

$$P(\text{“remember”}|\text{New}) = \Phi\left(\frac{-C_o}{s}\right)\Phi\left[\frac{1}{s}\left(\frac{d_x^2 - d_y^2}{\sqrt{d_x^2 + d_y^2}} + C_r\right)\right]. \tag{B3}$$

The remember hit rate is likewise a fraction of the hit rate:

$$P(\text{“remember”}|\text{Old}) = \Phi\left(\frac{2d_x d_y}{\sqrt{d_x^2 + d_y^2}} - C_o\right)\Phi(C_r). \tag{B4}$$

Equations B1–B4 have four free parameters: the diagnosticity of global and specific information (d_x and d_y), the old–new criterion location relative to the mean of the New distribution (C_o), and the R-K criterion location relative to the mean of the Old distribution (C_r). The fifth parameter, s, is fixed at 0.8 in our simulations of standard R-K and R-K rating experiments but is estimated as the slope of the zROC curve in O-N rating experiments. The results of a single experimental task (an old–new judgment, old responses followed by an R-K decision) are thus sufficient to estimate the parameters.

Received May 1, 2001
Revision received July 28, 2003
Accepted July 31, 2003 ■