Psychological Review
2006, Vol. 113, No. 3, 648–656

0033-295X/06/$12.00 DOI: 10.1037/0033-295X.113.3.648

COMMENTS

Decision-Making Models of Remember–Know Judgments: Comment on Rotello, Macmillan, and Reeder (2004)

Bennet Murdock
University of Toronto

The sum-difference theory of remembering and knowing (STREAK) provides a sophisticated account of many interactions in the remember–know (R–K) area (C. M. Rotello, N. A. Macmillan, & J. A. Reeder, 2004). It assumes 2 orthogonal strength dimensions and oblique criterion planes. Another dual-process model (J. T. Wixted & V. Stretch, 2004) with one decision axis has also been applied to R–K judgments with considerable success and provides new insights into the processes involved. An analysis of the 4 major R–K interactions can also be explained by a simpler one-dimensional signal detection theory (J. C. Dunn, 2004a). However these models do not make contact with standard work on recognition memory, so their scope is limited. To bridge this gap, a global-matching model (a theory of distributed associative memory [TODAM]) for R–K judgments is proposed. This model can produce good fits to the data, and there are established experimental manipulations with which to test it. It provides further support for the idea that R judgments are based on associative information, whereas K judgments are based on item information.

Keywords: remember–know, STREAK, TODAM, models

Several decision-making models of remember (R) and know (K) judgments have recently appeared. These are the sum-difference theory of remembering and knowing (STREAK; Rotello et al., 2004), a unidimensional dual-process signal-detection model (Wixted & Stretch, 2004), and a signal-detection analysis of R–K interactions supporting a one-dimensional interpretation (Dunn, 2004a). In this article, I comment on these models and point out that none of them make contact with more traditional empirical and theoretical work on recognition memory. I close by suggesting that an extension of TODAM, which is based on item information and associative information, is simpler than STREAK, can explain the Dunn interactions, and provides a principled account of the memory processes that might underlie R–K judgments.

STREAK

STREAK provides an impressive account of a large amount of R–K data. It assumes two orthogonal strength dimensions (*global strength* and *specific strength*) and two orthogonal criterion that are oblique with respect to the strength axes. The model is faithful to the original distinction between R and K judgments (Tulving, 1985), which suggests that they reflect the contribution of familiarity and recollection, assumed to be two different forms of memory. On the other hand, STREAK contrasts with many other R–K models that assume one strength dimension with two decision criterion (Donaldson, 1996).

According to STREAK, in an R–K experiment the subject first has to decide whether the sum of global and specific strength values is above or below an old–new criterion plane that is oblique (oriented southeast to northwest) relative to the *x*, *y*-axes. If this sum is above the old–new criterion plane, then the subject must make an R or K judgment. The subject gives a *remember* response if the difference between the global and specific strength is above (northwest of) an R–K criterion plane orthogonal to the old–new criterion plane but gives a *know* response if this difference is below (southeast of) this criterion plane. If the first test fails (the sum of the global and specific strengths is below the old–new criterion), then a *new* response results.

Rotello et al.'s (2004) primary test of the one-dimensional signal-detection model was to fit a large number of R–K experiments and then use those results to generate two-point receiver operating characteristic (ROC) slopes, which they found generally to have a slope of one. This led to the development of STREAK. They developed expressions for the STREAK model so they could fit these two-point ROC curves in one step, but this is a computational device not an aid to understanding.

Although the fits to the data were good, Rotello et al. (2004) noted that the model was "saturated." They used four parameters (the bivariate means of the old-item strength distributions and the offsets [distance] of the two criterion planes from the means of the old- and new-item distributions) and then used these four param-

Preparation of this article was supported by Research Grant APA146 from the Natural Sciences and Engineering Council of Canada. I thank Matthew Duncan, Bill Hockley, Mike Kahana, Ken Malmberg, Bruce Oddson, and Dave Smith for their helpful comments.

Correspondence concerning this article should be addressed to Bennet Murdock, Department of Psychology, University of Toronto, Toronto, Ontario, M5S 3G3 Canada. E-mail: murdock@psych.utoronto.ca

eters to generate the four points necessary for the two-point ROC curves.

The fact that the model was saturated means that the good fits cannot be considered strong support for the model. Also, one can question the identification of the two strength axes. They are described in the Rotello et al. (2004) article as global strength and specific strength (see their Figure 4, p. 592). It seems unlikely that global and specific strengths are orthogonal (independent); they could well be oblique (correlated). Further, it is even less clear how one could vary specific and global similarity independently.

As one of the architects of the STREAK model has suggested,[1] it could be argued that it is not important what the strength axes are called as long as the decision criteria are specified. However, if the nature of the strength axes is unknown, the model cannot be tested by varying the values on the two dimensions and, as I discuss later in the article, this may be an important way to test any of the R–K models.

STREAK claims that the slopes of R–K and *z*-transformed ROC curves differ, but it has been suggested that this depends on the method of analysis and in fact may not be useful in distinguishing between models (Malmberg & Xu, in press). However, in rebuttal, this suggestion has also been challenged (Rotello, Macmillan, Hicks, & Hautus, in press).

There may be some confusion about the sum and difference point in STREAK. In particular, are decisions in STREAK made on the basis of the values of x and y (global and local strength) or on sums and differences ($x + y$ and $x -$ y)? The verbal descriptions in the model suggest the latter, but Rotello et al.'s (2004) Figures 4 and 5 (pp. 592, 594) suggest the former. There are two possible interpretations. The first interpretation is that the decisions are based on the values of x and y, but the results of the decisions reflect sums and differences. The second possible interpretation is that the decisions are indeed based on sums and differences.

From a process point of view, decisions can be thought of as being a stage of processing in which there is an input, a (decision) function, and an output. The decision function is greater than or less than, and the input is either x and y (first interpretation) or sums and differences (second interpretation). In both cases, the output depends on the input and the result of the decision function operating on the input, and perhaps the sum-and-difference interpretation reflects the output given the first interpretation.

Why does it matter which interpretation is correct? Because the results are quite different, the predicted values of R, K, and new differ accordingly. If one wants to simulate the model, the alternative to choose must be known. In the course of the peer review of this article, one of the reviewers argued that simulations are not necessary; the derivations give the results for any specified parameter values. However, my feeling is that the model should be simulated in order to check on one's understanding. Because the integral equations are not provided, one must rely on one's own interpretation of a verbal description for guidance, and verbal descriptions can be unreliable. As an example of what I mean by integral equations see Equations 3–6 presented subsequently.

I am not saying that the derivations are wrong. What I am saying is that these derivations must be available for public scrutiny, and without the integral equations one must rely on simulations to check both one's understanding of the model and the accuracy of the model derivations. This way, any possible confusion about the model should not arise.

I have tried to simulate the model but not completely successfully. However, what is very clear is that for several selected data sets that use the STREAK parameter values, the predicted results are much closer to the obtained results for the first interpretation than for the second interpretation. So perhaps the sum-and-difference interpretation suggested by Rotello et al. (2004) really applies to the output of the decision function rather than to the input. Then, as Rotello et al. have suggested, old–new judgments reflect more the sum of global and local strengths, whereas R–K judgments reflect more their difference; but this is the output not the input of the decision process.

Because the criteria are oblique relative to the axes, for a subject to make decisions (or to simulate the model) the subject must specify what these functions are and make the decision on the basis of the functions not the arguments of the functions. If μx and μy (d_x and d_y in the Rotello et al. [2004] article) are the means of the bivariate old-item distribution on the x (global strength) and y (specific strength) dimensions, the slopes of these two orthogonal criterion functions are $-\mu y/\mu x$ and $\mu x/\mu y$ for the old–new and R–K judgments, respectively. Although these slopes are given in the Rotello et al. article the intercepts are not. The intercepts are given below for readers who might wish to simulate the model.[2] (C_o and C_r are the offsets of the old–new and R–K criterion cuts from the means of the new- and old-item distributions, respectively.)

$$\text{int old} = \text{new} = \frac{Co\sqrt{\mu x^2 + \mu y^2}}{\mu x} \tag{1}$$

$$\text{int R–K} = \frac{Cr\sqrt{\mu x^2 + \mu y^2} - \mu y^2 + \mu x^2}{-\mu y} \tag{2}$$

It should also be mentioned that there are two procedures used for R–K experiments. In one, call it the ONRK procedure, subjects first give an *old* or *new* response, then, if *old*, they go on to make an R–K response. In the other, call it the RKN procedure, subjects first make the R–K judgment then, if neither R nor K, they go on to make a *new* response. As I understand the STREAK model, it would not make any difference which procedure is used; the same parameter-estimation procedure would be used regardless. This assumes that subjects in the RKN procedure implicitly decide whether the probe item is old or new, and they would not make an R–K decision if the first decision was *new*. Both methods were used in the experiments reported in the Rotello et al. (2004; Appendix A) article (far more of these experiments used the ONRK than the RKN procedure), but apparently the same parameter-estimation procedure was used in both cases.

A Unidimensional Dual-Process Signal-Detection Model

Another dual-process model with a single decision axis has been proposed and also shown to be consistent with many of the standard R–K results (Wixted & Stretch, 2004). In particular, it shows how R and K can vary qualitatively and explains the fact

[1] Neil Macmillan made this point in a talk at the University of Toronto.

[2] I thank Caren Rotello for providing these.

that data from R–K experiments often fall on the same normalized ROC (z-ROC) curve as points from a confidence-judgment procedure. This dual-process model assumes that, though there are dual processes (recollection and familiarity), there is only a single dimension for the decision axis. This dimension is the same strength axis as in classical signal detection theory.

This unidimensional dual-process signal-detection model is illustrated by assuming a normally distributed preexperimental familiarity for new items and added strength in the form of increased familiarity and increased recollective strength, both rectangularly distributed. Its main strength is claimed to be that it can explain many findings such as R and K false alarms so there are criterion effects for both *remember* and *know* responses. Although this is clearly a strong argument for this approach, recollection and familiarity are the results of processes not processes themselves. Also, the assumption that recollection and familiarity are the underlying distributions seems to presuppose that which needs to be explained.

One-Dimensional Models

After an extensive review of the R–K literature, Dunn (2004a) claimed that the original one-dimensional signal detection theory (TSD) model of R–K findings (Donaldson, 1996) can explain the four basic interactions that have seemed problematic for such a model. Dunn (2004b) also developed a strong test that has the potential to reject all one-dimensional TSD models, and, in an extensive review of the literature, he failed to find any results that would reject a one-dimensional view. Smith and Duncan (2004) tested theories of recognition by predicting performance across paradigms and showed that the fits to a one-dimensional TSD model are better than the fits to a particular two-dimensional model. However, Smith and Duncan argued that these fits are "constrained." They seem to feel that a dual-process model is more likely.

Making Contact With Recognition Memory Models

Although these are all impressive and potentially useful models, there is a general issue that applies to all of them. They are all restricted to a particular paradigm (the R–K paradigm) and fail to make contact with many other areas of recognition memory. They do not address such basic issues as serial-position and presentation rate effects (Strong, 1912, 1913), word-frequency effects (Shepard, 1967), the spacing effect (Hintzman, 1974), linear reaction time functions for subspan lists (Sternberg, 1966) and for supraspan lists (Murdock & Anderson, 1975), the mirror effect (Glanzer & Adams, 1985), and the list-strength effect (Ratcliff, Clark, & Shiffrin, 1990).

These are the issues discussed by the various global matching models such as the composite holographic associative recall model (CHARM; Eich, 1982), TODAM (Murdock, 1982), the search of associative memory or SAM model (Gillund & Shiffrin, 1984), MINERVA (Hintzman, 1988), the target–episode–cue–object or TECO model (Sikstrom, 1996), the retrieving effectively from memory or REM model (Shiffrin & Steyvers, 1997), the subjective-likelihood model (McClelland & Chappell, 1998), the bind–cue–decide or BCDmem model (Dennis & Humphreys, 2001), the dual-process source of activation confusions or SAC model (Reder et al., 2000), and a model similar to SAC that deals with R–K judgments by using distributed memory mechanisms (Norman & O'Reilly, 2003).

On the other hand, most of the global matching models do not address the R–K data, but there are exceptions. First, the BCDmem model deals with data from experiments on the process-dissociation procedure (Jacoby, 1991), and this was a forerunner to the R–K work. Second, the SAC model cannot only deal with the R–K data, but it can also explain in detail interactions with word frequency over repetitions. Third, an application of REM to R–K judgments is implied by a study of midazolam effects on episodic recognition memory (Malmberg, Zeelenberg, & Shiffrin, 2004). However, even these applications do not deal in detail with the R–K interactions discussed in STREAK, the Dunn (2004a) article, or the Wixted and Stretch (2004) analysis.

Bridging the Gap

The strength of a signal-detection approach is that it provides a way of separating memory and decision, but it is not a process approach and it does not explain the basis for the assumed underlying strength or what factors should affect it. The strength of the STREAK and Wixted and Stretch (2004) models is that they provide deeper insights into R–K data but, as noted, they do not make contact with standard work on recognition memory (empirical and theoretical) and do not give us any clear guidelines as to what the basis for the assumed "strengths" is or how they might be modified. More generally, STREAK, the Wixted and Stretch account, and the Donaldson and Dunn signal-detection account are all incomplete in that they do not provide any theoretical account of the processes they take for granted. To provide a more theoretically based account, I suggest a slight extension of TODAM to account for R–K judgments. This extension is similar to the dual-process model of Wixted and Stretch, fits the interactions of Dunn and the R–K confidence-judgment ROC plots, has a simpler decision mechanism than STREAK because the decision criteria are orthogonal not oblique, and makes contact with the extensive data on item recognition mentioned above.

TODAM

In TODAM items are represented by random vectors (Anderson, 1970), that is, vectors of features sampled from zero-centered normal distributions. Associations between two items are represented by convolution, retrieval is represented by correlation (Borsellino & Poggio, 1973), and both items and associations are stored in a common memory vector (Murdock, 1982). For item recognition, the probe item is compared with the memory vector by taking the dot product of the two, and, for recall, the probe is correlated with the memory vector that generates a noisy approximation to the target item that must be deblurred (Liepa, 1977).

TODAM has always assumed that, when a list of paired associates is presented, the subject encodes (stores) both item information and associative information. A recent study of the use of different types of associative information in modeling recognition and recall (Kahana, Rizzuto, & Schneider, 2005) discusses item information, autoassociative information, and heteroassociative information. To apply this item-associative framework to R–K judgments, assume that the same is true in studies of recognition and R–K judgments. Then we would have two orthogonal dimensions, but they are not global and specific strength, rather, as in SAC and the Yonelinas model (Yonelinas, 1997), they are *item information* and *associative information*. The item information is the strength or resonance (Ratcliff, 1978) of the dot product of the probe vector with the memory vector, and the associative information is the resemblance (also strength) of the retrieved item to its representation in the memory vector, again as measured by the dot product.

For an RKN procedure, assume that subjects first assess the strength of the associative information, and if the associative strength is above the associative criterion give a *remember* response. If it is not (i.e., the strength of the associative information is below the R criterion), assess the strength of the item information and give a *know* response if the item strength is above the K criterion. If the strength of the item information is below the K criterion, give a *new* response. The R and K criteria are separate and distinct, so the question of where the item observation is relative to the R criterion does not arise.

The idea of two distinct steps has been suggested before (Atkinson & Juola, 1974; Mandler, 1980) so this is nothing new. This TODAM model is a dual-process model in the true sense; there are two successive processes (doing the correlation–comparison and taking the dot product), the results of which (resemblance of the retrieved item to the probe itself) form the basis of decision. For autoassociation, as in CHARM, the retrieved item would only have to be compared with the probe item to yield a resemblance value, whereas for cross-association, as in TODAM, the retrieved item would have to be compared with the memory vector to determine whether it was a good match to another list item. However, in either case (resonance and resemblance) there would only be a single decision axis (strength).

This analysis assumes that, during list presentation for the associative information, the subject either associates the item with itself, as in CHARM, or associates it with another list item, as in TODAM. This association is in addition to storing the item itself directly in the memory vector, and the direct storage constitutes the item information. Dual storage of item information and associative information in a single memory vector is also nothing new; it was part of the original version of TODAM (Murdock, 1982). In support of this approach, a recent study (Schwartz, Howard, Jing, & Kahana, 2005) showed that recollection was better for subsequently tested items that were presented near in time to the original presentation of the recollected item. Schwartz et al. (2005) stated that "recollection of an item not only retrieves detailed information about the item tested, but also retrieves information about the item's neighbors" (p. 901) (i.e., associative information).

Thus, in this TODAM model there are two dimensions (item and associative strength) that serve as the basis for decision, and there are two criteria. There are two processes (the correlation–comparison process and taking the dot product) and these are conditional. The second process (taking the dot product) only occurs if the first process (correlation and comparison) fails. The processes, of course, are identical for old and new items, so the three possible responses for old and new items are *remember*, *know*, and *new*. The latencies of *remember* responses should be faster than the latencies of *know* responses (because the second process only occurs if the first process fails), and perhaps *know* responses should be faster than *new* responses because, in studies of item recognition, *old* responses are generally faster than are *new*. *Remember* responses are clearly faster than *know* responses (Dewhurst & Conway, 1994; Hockley, Hemsworth, & Consoli, 1999), but *new* responses are not always the slowest. Also a study of retention-interval effects found that hits were faster than *know* at all lags and misses were intermediate, though closer to the *know* than to the *remember* judgments (Rubin, Hinton, & Wenzel, 1999, Figure 12, p. 1171).

In contrast to the STREAK model, item and associative information are demonstrably independent. This is shown for autoassociations (associating the item with itself) in the present Appendix, and the same is true when the items are associated with other list items. (They could even be associated with nonlist items if context were included in the item representations.) In a detailed analysis of correlations in recognition and recall, it has been shown that TODAM, CHARM, and the matrix model can all fit findings on the correlation between item recognition and cued recall in the successive testing procedure (Kahana et al., 2005).

On each dimension (item information and associative information), the criterion dimension is orthogonal to the respective strength. Subjects do not need to make estimates of the means of an old-item bivariate distribution to set their criteria. With the standard TODAM assumptions, both item information and associative information are approximately normally distributed so typical ROC curves would be expected. Finally, it is clear how to manipulate item and associative information independently to test the model (Murdock, 1997).

To show that this model can fit the data, let f_O and f_N be the old and new associative-information distributions, and let g_O and g_N be the old and new item-information distributions. Let *a* be associative criterion or the criterion cut on the associative information dimension that must be exceeded for a *remember* response, and let *b* be the item criterion or the criterion cut on the item information dimension that must be exceeded for a *know* response. Then if R is the probability of a *remember* response, K is the probability of a *know* response, O is an old item and N is a new item and *s* is strength:

$$R_O = \int_{s=a}^{\infty} f_O(s)ds \tag{3}$$

$$K_O = (1 - R_O)\int_{s=b}^{\infty} g_O(s)ds \tag{4}$$

$$R_N = \int_{s=a}^{\infty} f_N(s)ds \tag{5}$$

$$K_N = (1 - R_N)\int_{s=b}^{\infty} g_N(s)ds \tag{6}$$

To fit any set of data, assume all the distributions (f_O, f_N, g_O and g_N) have unit variances and the two new-item distributions (f_N and g_N) have mean zero. (The assumption of unit variance for the two old-item distributions is discussed later.) Then we have four parameters we must solve to fit any set of R–K data; namely, $\mu(f_O)$, $\mu(g_O)$, a, and b where $\mu(f_O)$ the associative mean and $\mu(g_O)$ the item mean are the means of the associative information and item information old-item distributions. If $\Phi(z)$ denotes the integral (area) of a unit normal curve between $-\infty$ and z, and $\Phi^{-1}(p)$ gives the z score corresponding to the probability p or the area between $-\infty$ and z, we can find the four parameter values by solving the following set of four equations. Given that R_N and K_N are the R and K response rates for new items then to find the estimates of criteria a and b denoted $\hat{a}$ and $\hat{b}$, then as (see Equation 5) $R_N = 1 - \Phi(a)$, the predicted value of a or $\hat{a}$ is

$$\hat{a} = \Phi^{-1}(1 - R_N). \tag{7}$$

Then as (see Equation 6) $K_N = (1 - R_N)(1 - \Phi[b])$, the predicted value of b or $\hat{b}$ is

$$\hat{b} = \Phi^{-1}\left(1 - \frac{K_N}{1 - R_N}\right). \tag{8}$$

Given that R_O and K_O are the R and K response rates for old items, then to estimate the parameters denoted as $\hat{\mu}(f_O)$ and $\hat{\mu}(g_O)$ the following equations are used. Because (see Equation 3) $R_O = 1 - \Phi(a—m\,[f_O]/s\,[f_O])$ as $\sigma(f_O) = 1$, the predicted value of $\mu(f_O)$ or $\hat{\mu}(f_O)$ is

$$\hat{\mu}(f_O) = \hat{a} - \Phi^{-1}(1 - R_O); \tag{9}$$

Because (see Equation 4) $K_O = (1 - R_O)(1 - \Phi[b - \mu\{g_O\}/\sigma\{g_O\}])$ as $\sigma(g_O) = 1$, the predicted value of $\mu(g_O)$ or $\hat{\mu}(g_O)$ is

$$\hat{\mu}(g_O) = \hat{b} - \Phi^{-1}\left(1 - \frac{K_O}{1 - R_O}\right). \tag{10}$$

As a numerical example, if $R_N = .618$, $K_N = .221$, $R_O = .945$, and $K_N = .040$, then (see Equation 7) $\hat{a} = -0.3$, (see Equation 8) $\hat{b} = -0.2$, (see Equation 9) $\hat{\mu}(f_O) = 1.3$, and (see Equation 10) $\hat{\mu}(g_O) = 0.4$.[3]

In his article, Dunn (2004a) analyzed four sets of experiments (Gardiner & Java, 1990; Gardiner, Kaminska, Dixon, & Java, 1996; Gregg & Gardiner, 1994; Schacter, Verfaellie, & Ames, 1997). These studies were selected because they represent all possible interactions of a 2 × 2 design. I found the values of $\hat{a}$ and $\hat{b}$ and $\hat{\mu}(f_O)$ and $\hat{\mu}(g_O)$ obtained from the R and K values in Dunn's Table 1 (columns 4 and 5, p. 527) by using Equations 7–10 to find the proportion of *know* and *remember* responses to old and new items predicted by the TODAM model. The results of this analysis are shown here in Table 1.

One might wonder how, in TODAM, as the item vectors are normalized to 1, the means of the R and K distributions could be greater than 1. However, TODAM is a global-matching model and the probe item is compared with the memory vector that contains all the items. Given any intralist similarity greater than zero as well as context, the dot product would reflect these factors.

As a check, I simulated the TODAM model by using the estimates of the four parameter values shown in Table 1 to predict R_O, K_O, R_N, and K_N and, as would be expected, in all cases the predicted values for the obtained values fit perfectly. There was a discrepancy in the 1 trial–4 trial value between Dunn's (2004a) Table 1 and Dunn's Excel database, but I used the former as they are more readily available.

Table 1
Estimated Criterion Parameters ($\hat{a}$ and $\hat{b}$) and Means ($\hat{\mu}$) for the Old Associative Information (f_O) and Item Information (g_O) Distributions for Four Experiments, Each With Two Conditions

Condition	$\hat{a}$	$\hat{b}$	$\hat{\mu}$ (f_O)	$\hat{\mu}$ (g_O)
Schacter et al., 1997 (Experiment 1)				
Amnesic	1.55	0.91	0.51	0.50
Control	1.88	1.39	1.75	1.41
Gregg & Gardiner, 1994 (Experiment 2)				
Auditory	1.64	1.31	0.36	0.79
Visual	1.88	0.89	0.65	1.10
Gardiner & Java, 1990 (Experiment 2)				
Word	1.75	1.20	1.17	0.44
Nonword	1.88	1.16	1.00	0.83
Gardiner et al., 1996 (Experiments 1 & 2)				
1 Trial	1.23	1.01	0.65	0.84
4 Trials	2.05	1.33	1.74	1.84

The point about perfect fits is discussed shortly, but first compare this analysis with the conclusion from the Dunn (2004a) analysis. In the first set of data (Gardiner & Java, 1990), the Dunn analysis gives a d' 1.85 for the normal subjects (controls) and 0.60 for the amnesic subjects and higher thresholds for the normal subjects than for the amnesic subjects for both *remember* (1.98 to 1.62) and *know* (1.25 to 0.75) judgments. The TODAM R–K analysis concurs with the general conclusion but in addition shows higher means and criteria for associative information than for item information (*M*s: 1.75–1.41; Criteria: 1.88–1.39) for the normal subjects but only higher criteria for associative information than for item information (1.55–0.91) for the amnesic subjects; the means for associative information and item information were the same (0.51–0.50).

Thus, the TODAM R–K model gives a more detailed account of the data but of course at the cost of one more degree of freedom (*df*). The *df* advantage for the Dunn (2004a) analysis is important because, unlike the TODAM, STREAK, and Wixted models, the Dunn analysis is in principle falsifiable by a statistical test, whereas these others are not. On the other hand, all these other models are in principle falsifiable by experimental tests and, because they give a deeper analysis of the processes underlying R–K judgments, they are probably more likely than the Dunn analysis to suggest such tests.

What about the other interactions? For the second set of data (an auditory–visual comparison reported by Gregg & Gardiner [1994]), the Dunn analysis shows a d' advantage for the visual over the auditory, but the R–K criteria interact with modality. Visual has a higher R criterion than does the auditory, but the reverse is true for the visual (visual: 2.43–2.12; auditory: 0.88–1.19). Exactly the same pattern is found for the TODAM model; *M*s = 0.65–0.36, visual to auditory for associative information; *M*s = 1.10–0.79, visual to auditory for item information; whereas,

[3] This method was suggested by John Dunn, and I appreciate this contribution.

the criteria are 1.88–1.64, visual to auditory for associative information but are 0.89–1.31, auditory to visual for item information.

For the third set of data (a comparison of words vs. nonwords reported by Gregg & Gardiner, 1994) the Dunn (2004a) analysis shows most of the effect is in the remember criteria. The d' estimates are similar (1.07–1.04) for words to nonwords as are the K criteria (1.07–1.04), but the R criteria are appreciably higher for the nonwords (1.89) compared with for the words (1.55). Note that in the data, all the effects are in the words (old words have a higher remember rate and a lower know rate than do new words, whereas in the new words remember and know rates are virtually identical [.01 difference] for words and nonwords). For TODAM, the picture is quite different. The associative information mean is slightly higher (1.17–1.00) for words than nonwords but the z score is almost twice as high (0.83–0.44) for nonwords than for words. If the words and nonwords are considered high- and low-frequency, as item information probably mediates simple recognition, the know-rate difference of the TODAM analysis seems consistent with more general recognition results than does the Dunn (2004a) analysis.

For the fourth set of data (a comparison of one trial vs. four trials reported by Schacter et al., 1997) the Dunn (2004a) analysis shows that the d' estimate is much higher (2.25–0.87) for four trials than for one trial, and this must override the criteria that are in the same direction for both R (2.56–1.45) and K (1.30–0.70) four trials to one trial. Again, the TODAM analysis concurs but carries the analysis one step further. Both associative information and item information means are higher for four trials than for one trial (associative information: 1.74–0.65; item information: 1.93–0.84) and, as in Dunn, so are the estimates for the criteria a and b, which are 2.05–1.23 and 1.33–1.01 for four trials versus one trial. Thus there is some agreement (notably the second and fourth data set) but some disagreement, and the latter could provide a basis for further testing.

Unlike STREAK, in this TODAM model no grid search is necessary to find the best fitting parameters. All it takes is a table of the normal curve and a hand calculator. The model will always fit perfectly within rounding errors so a good fit is not a test of the model. However the parameter estimates suggest what processes might underlie R and K judgments according to the TODAM R–K model.

How could the model explain slopes of ROC curves <1? There are several possible ways. One is by including context although that might pose problems for the TODAM interpretation of the list-strength effect which used the continuous memory assumption (Murdock & Kahana, 1993). Another would involve the notion of probabilistic encoding. According to TODAM, only some features are added to the memory vector with every presentation, and if the distribution of that probability is variable across subjects and items, then TODAM can account for the slope of the z-ROC being <1 (Kahana et al., 2005).

The TODAM model suggested here applies to experiments that use the RKN procedure. What about experiments in which an ONRK procedure is used? Perhaps as in STREAK the same logic would work for both procedures, or perhaps the flowchart of the TODAM model would have to be slightly different to accommodate these findings. Probably no new principles would be needed, but this is a problem for further development. It might be mentioned that three of the four experiments discussed by Dunn (2004a) used an RKN procedure.

One might object to the simplicity argument because convolution is not simple. It may not be familiar, but it is not complex. Connectionist models routinely use outer-product matrices to represent associative information, and the convolution–correlation formalism is only one additional step beyond an outer-product matrix. It can be represented as summing the antidiagonals (convolution) and diagonals (correlation) of an outer-product matrix, and these summations turn a matrix into a vector. It was proposed long ago as a possible formalism for storage and retrieval in human memory (Borsellino & Poggio, 1973) and has always been used in TODAM.

As noted above, the TODAM model cannot fail. This may seem surprising to some readers, but in fact exactly the same thing is true of the two-process one-dimensional signal-detection model of Wixted and Stretch (2004; apparently, this is not true of STREAK).[4] More generally, this is true of signal detection theory in general; you can always find a set of parameters (originally d' and β) that fit any set of 2×2 recognition-memory data perfectly. This comment does not apply to the Dunn (2004a) fits of the R–K interactions because there are three parameters and four *df*s.

Does this mean all these theories are worthless? Of course not. Each model or theory presents a unique description of the data that provides another way of understanding it and, thus, leads to different ways of testing it. The original TSD model provided a novel way of separating memory and decision, and that is testable; finding the parameter values by themselves is not the contribution. The contribution is finding out, through testing the model, how the parameter values change given the experimental manipulations and whether these changes agree with the predicted results.

The TODAM model suggested here can easily be tested. Vary the relative strengths of item information and associative information and determine whether the parameters behave accordingly. This was done for two Item Information $\times$ Associative Information interactions (Hockley, 1992; Hockley & Cristi, 1996). The results showed clearly that the original version of TODAM was flawed and led to TODAM2 (Murdock, 1997).

The good fits to the interactions reported by Dunn (2004a) and replicated here should illustrate the fact that finding interactions between experimental conditions is not going to settle the one-process versus two-process issue. This is why it is important to go beyond fitting the models to data and determining whether the assumed mechanisms have the predicted effect (Roberts & Pashler, 2000). For R–K data, the fit is guaranteed; the test is whether the parameter values vary as predicted or the predictions are supported by experimental evidence.

Quite apart from the strength and weaknesses of the various models discussed here, it is important to realize that fitting these

[4] In a personal correspondence, Caren Rotello wrote that the published version of STREAK was developed only after simpler models failed. As to the point about saturated models failing, there is a common misunderstanding on this point. Even a "saturated" process model can fail (to fit the data). The common view that one can fit any data set of N values with N free parameters applies to polynomial models not to process models. I once was unable to fit a 7-point serial-position curve with 15 free parameters.

models to data, even explaining interactions, is not much of a test. Or, it is a necessary but not a sufficient condition. It is probably not an exaggeration to say that essentially all the models can fit the data (at least the interactions) perfectly. We must take the next step and show that experimental manipulations specified by the models have the intended effect. We should use the parameter values obtained not as an index of goodness of fit but rather to see whether the results move in the directions predicted by the model given the appropriate experimental manipulations.

How do we do this? For each model, we must explore the parameter space and derive results or simulate them with a variety of parameter values. Then the model can be better understood and, hopefully, experimental manipulations that should discriminate among the models can be found. Not only will the database be enriched, but also the understanding of the processes involved will be deeper and broader.

Conclusions

Four general conclusions are suggested. First, in the interests of cumulative development of knowledge, models of R–K should build on, or at least make contact with, existing theories of memory rather than operating as if a whole literature on memory did not exist or was irrelevant. Second, processes that are postulated to determine R or K responses should be given a substantive meaning within a theoretical framework rather than just being convenient labels attached to the outputs of such processes. Third, given explicit definitions of the processes involved, it should be possible to identify experimental factors that affect one or the other process, thereby rendering the enterprise falsifiable. Fourth, an extension of TODAM to R–K judgments can fit the data, including the Dunn (2004a) interactions, and adds support to previous suggestions that R–K judgments are based on associative information and item information, respectively.

References

Anderson, J. A. (1970). Two models for memory organization using interacting traces. *Mathematical Biosciences, 8,* 137–160.

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 developments in mathematical psychology* (Vol. 1, pp. 243–293). San Francisco: Freeman.

Borsellino, A., & Poggio, T. (1973). Convolution and correlation algebras. *Kybernetik, 122,* 113–122.

Dennis, S., & Humphreys, M. S. (2001). A context noise model of episodic word recognition. *Psychological Review, 108,* 452–478.

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

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

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

Dunn, J. C. (2004b, July). *A strong inference test of the signal-detection interpretation of remember–know judgments.* Paper presented at the 3rd Annual Summer Interdisciplinary Conference, Cavalese, Italy.

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

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

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

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

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

Gregg, V. H., & Gardiner, J. H. (1994). Recognition memory and awareness: A large effect of study–test modalities on "know" responses following a highly perceptual orienting task. *European Journal of Cognitive Psychology, 6,* 131–147.

Hintzman, D. L. (1974). Theoretical implications of the spacing effect. In R. L. Solso (Ed.), *Theories in cognitive psychology* (pp. 77–99). Hillsdale, NJ: Erlbaum.

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

Hockley, W. E. (1992). Item versus associative information: Further comparisons of forgetting rates. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 18,* 1321–1330.

Hockley, W. E., & Cristi, C. (1996). Tests of encoding tradeoffs between item and associative information. *Memory & Cognition, 24,* 202–216.

Hockley, W. E., Hemsworth, D. H., & Consoli, A. (1999). Shades of the mirror effect: Recognition of faces with and without sunglasses. *Memory & Cognition, 27,* 128–138.

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

Kahana, M. J., Rizzuto, D. S., & Schneider, A. R. (2005). Theoretical correlations and measured correlations: Relating recognition and recall in four distributed memory models. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 31,* 933–953.

Liepa, P. (1977). *Models of content addressable distributed associative memory (CADAM).* Unpublished manuscript, University of Toronto, Toronto, Ontario, Canada.

Malmberg, K. J., & Xu, J. (in press). The influence of averaging and noisy decision strategies on the recognition memory ROC. *Psychonomic Bulletin & Review.*

Malmberg, K. J., Zeelenberg, R., & Shiffrin, R. M. (2004). Turning up the noise or turning down the volume? On the nature of the impairment of episodic recognition memory by midazolam. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 30,* 540–549.

Mandler, G. (1980). Recognizing: The judgment of previous occurrence. *Psychological Review, 87,* 252–271.

McClelland, J. L., & Chappell, M. (1998). Familiarity breeds differentiation: A subjective-likelihood approach to the effects of experience in recognition memory. *Psychological Review, 105,* 734–760.

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

Murdock, B. B. (1997). Context and mediators in a theory of distributed associative memory (TODAM2). *Psychological Review, 104,* 839–862.

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

Murdock, B. B., & Kahana, M. J. (1993). An analysis of the list-strength effect. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 19,* 689–697.

Norman, K. A., & O'Reilly, R. C. (2003). Modeling hippocampal and neocortical contributions to recognition memory: A complementary-learning-systems approach. *Psychological Review, 110,* 611–648.

Ratcliff, R. (1978). A theory of memory retrieval. *Psychological Review, 85,* 59–108.
Ratcliff, R., Clark, S. E., & Shiffrin, R. M. (1990). The list-strength effect: I. Data and discussion. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 16,* 163–178.
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 computational model of remember–know judgments in a continuous recognition paradigm. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 26,* 294–320.
Roberts, S., & Pashler, H. (2000). How persuasive is a good fit? A comment on theory testing. *Psychological Review, 107,* 358–367.
Rotello, C. M., Macmillan, N. A., & Reeder, J. A. (2004). Sum-difference theory of remembering and knowing: A two-dimensional signal-detection model. *Psychological Review, 111,* 588–616.
Rotello, C. N., Macmillan, N. A., Hicks, J. L., & Hautus, N. J. (in press). Interpreting the effects of response bias on remember-know judgments using signal-detection and threshold models. *Memory & Cognition.*
Rubin, D. C., Hinton, S., & Wenzel, A. (1999). The precise time course of retention. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 25,* 1161–1176.
Schacter, D. L., Verfaellie, M., & Ames, M. D. (1997). Illusory memories in amnesic patients: Conceptual and perceptual false recognition. *Neuropsychology, 11,* 331–342.
Schwartz, G., Howard, M. C., Jing, B., & Kahana, M. J. (2005). Shadows of the past: Temporal retrieval effects in recognition memory. *Psychological Science, 16,* 898–904.
Shepard, R. N. (1967). Recognition memory for words, sentences, and pictures. *Journal of Verbal Learning and Verbal Behavior, 6,* 156–163.
Shiffrin, R. M., & Steyvers, M. (1997). A model for recognition memory: REM–retrieving effectively from memory. *Psychonomic Bulletin and Review, 4,* 145–166.
Sikstrom, S. (1996). The TECO connectionist theory of recognition failure. *European Journal of Cognitive Psychology, 8,* 341–380.
Smith, D. G., & Duncan, M. (2004). Testing theories of recognition memory by predicting performance across paradigms. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 30,* 615–625.
Sternberg, S. (1966). High-speed scanning in human memory. *Science, 153,* 652–654.
Strong, E. K. J. (1912). The effect of length of series upon recognition memory. *Psychological Review, 19,* 447–462.
Strong, E. K. J. (1913). The effect of time-interval upon recognition memory. *Psychological Review, 20,* 339–372.
Tulving, E. (1985). Memory and consciousness. *Canadian Psychology, 26,* 1–12.
Wixted, J. T., & Stretch, V. (2004). In defense of the signal-detection interpretation of remember/know judgments. *Psychonomic Bulletin & Review, 11,* 616–641.
Yonelinas, A. R. (1997). Recognition memory ROCs for item and associative information: The contribution of recollection and familiarity. *Memory & Cognition, 1997,* 747–763.

Appendix

Here I consider the worst case, autoconvolution (the association of the item with itself). If **f** is an item vector consisting of *N* elements, all of which are random samples from a normal distribution with mean 0 and variance $1/N$, and **g** is an "associative" vector consisting of the autoconvolution of **f** with itself, then if *X*, *Y*, and *Z* are normally distributed random variables, the expectation E of the dot product (·) of **f** and **g** is $E(\mathbf{f} \cdot \mathbf{g}) = f_1(N)\,E(X^3) + f_2(N)\,E(X^2Y) + f_3(N)\,E(XYZ)$.

Because $E(X^iY^jZ^k\ldots) = 0$, if any *i*, *j*, or *k*. . . is odd (the expectation of the product of random variables in the product of the expectations and, for a normal distribution, the expectation of any odd power is zero) then $E(\mathbf{f} \cdot \mathbf{g}) = 0$. The same result would hold if the association was between two different items (i.e., **f** and some other list item).

Received June 1, 2004
Revision received October 20, 2005
Accepted October 25, 2005 ■

Postscript: Reply to Macmillan and Rotello (2006)

Bennet Murdock
University of Toronto

The main point of my comment was to try to highlight the importance of relating work on remember–know judgments to more traditional work on recognition memory. The TODAM model I suggested was an example. In their reply, Macmillan and Rotello (2006) did not comment on this, so I do not know whether they agree.

I may have misunderstood the STREAK model. On the basis of Figures 4 and 5 in Rotello, Macmillan, and Reeder (2004), I assumed that, as is generally the case in signal-detection type models, the decision axes were the same as the memory axes (global and specific strength). I now realize that they may have assumed a change of basis. Perhaps global and specific strengths are the memory axes, but the criterion lines (planes) oblique to the memory axes form the basis vectors for the decision system. If so, then observations from the memory system must be mapped onto the decision space to permit remember–know (R–K) judgments. If this is correct, then it is quite reasonable to use sums and differences as the basis for decision. However, this is a rather different model. Not only is there a change of basis (a rotation of the memory vectors through an angle [θ]), but the basis vectors are offset by Co and Cr (the offsets of the old–new and R–K criterion cuts from the means of the new- and old-item distributions, respectively). Consequently we have a function (sums and differences), an offset, and a change of basis. The change of basis is frequently used in work on vector spaces (Murdoch, 1970), and the change is described by the so-called "β" matrix; namely,

$$\beta = \begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix}.$$

These three functions are order dependent, so the order in which these operations are done affects the outcome, but STREAK does