Psychological Review
2006, Vol. 113, No. 3, 657–665


0033-295X/06/$12.00 DOI: 10.1037/0033-295X.113.3.657

# Deciding About Decision Models of Remember and Know Judgments: A Reply to Murdock (2006)

Neil A. Macmillan and Caren M. Rotello
University of Massachusetts at Amherst

B. B. Murdock (2006) has interpreted remember–know data within a decision space defined by item and associative information, the fundamental variables in his general recognition memory model TODAM (B. B. Murdock, 1982). He has related parameters of this extended model to stimulus characteristics for several classic remember–know data sets. The authors show that this accomplishment is shared by both one- and two-dimensional signal-detection-based remember–know models (J. C. Dunn, 2004; C. M. Rotello, N. A. Macmillan, & J. A. Reeder, 2004; J. T. Wixted & V. Stretch, 2004), which can be represented in this same decision space and can be related to stimulus characteristics with similar success. Murdock claims that his model, unlike its competitors, is a process model; however, the process aspects of TODAM are not used in his application, and the decision aspects are identical to a previously proposed model. Murdock's claim that one competing model (STREAK; C. M. Rotello et al., 2004) is not fully specified is shown to be false. The new model is not superior to existing ones, but comparisons among the models to date are not definitive. The authors describe several strategies that might be applied to distinguish among them.

*Keywords:* remember–know, recognition memory, memory models, recollection, familiarity

When a subject in a memory experiment reports that a test item is "remembered" or "known" to have been on a study list, what psychological variables or processes are being tapped? What evidence can discriminate among competing answers to this question? Murdock (2006), the most recent theorist to address these issues, answered that item and associative information underlie performance and that the key evidence concerns relating decision constructs to stimulus variables. In this brief article, we first show that existing remember–know models form a family that draws on a single representation; each family member offers a distinct proposal about the basis for decisions within the same decision space. We show that Murdock's effort to connect remember–know experiments to an existing theory of recognition memory does not distinguish among the models. In fact, all existing models have both strengths and weaknesses; we conclude by proposing three strategies that have the potential to decide among them.

## A Family of Remember–Know Models

Tulving (1985) proposed the remember–know paradigm as a strategy for distinguishing noetic and autonoetic consciousness, mental states that rely primarily on semantic and episodic information, respectively. The relation he envisioned between these concepts is shown in Figure 1, taken from his article. Within a two-dimensional space, an individual's memory mode varies from heavily semantic [point ($a$, $z$)] to heavily episodic [point ($c$, $x$)]. Events above a "conversion threshold" are recognized as old, whereas those below it are not.

This two-dimensional structure is consistent with several current models of the remember–know paradigm. One of the details over which these models disagree is the appropriate names to be attached to the axes; we evade this question (for the moment) by referring to them as $x$ and $y$, letting $x$ correspond to Tulving's (1985) semantic dimension and $y$ to his episodic dimension.[1] We describe the three models in the order they were introduced, although in two cases a more elaborate version is the focus of current research.

### *The One-Dimensional Model*

Donaldson (1996) suggested that remembered and known items differ on a single strength dimension. Old and new items generate normal distributions differing only in mean along this axis, and two criteria partition the axis into response categories. Events above the upper criterion lead to "remember" responses, those between the criteria lead to know "responses," and those below both criteria lead to "new" responses. This conception seems at odds with Tulving's (1985) essential two-process idea, but Wixted and Stretch (2004) recently suggested how the two views could be reconciled: Allow the strength axis to be a sum of "familiarity" and "recollection." We identify these labels with the $x$ and $y$ axes in the

Neil A. Macmillan and Caren M. Rotello, Department of Psychology, University of Massachusetts at Amherst.

We are supported by Grant R01 MH60274 from the National Institutes of Health. We thank John Dunn and John Wixted for their helpful comments on this article and Ben Murdock for the opportunity to exchange ideas on this topic.

Correspondence concerning this article should be addressed to Neil A. Macmillan, Department of Psychology, University of Massachusetts, Amherst, MA 01003. E-mail: nam@psych.umass.edu

[1] We adopt the opposite arrangement from that in Figure 1 because one current model (STREAK; Rotello et al., 2004) has already reversed the axes.

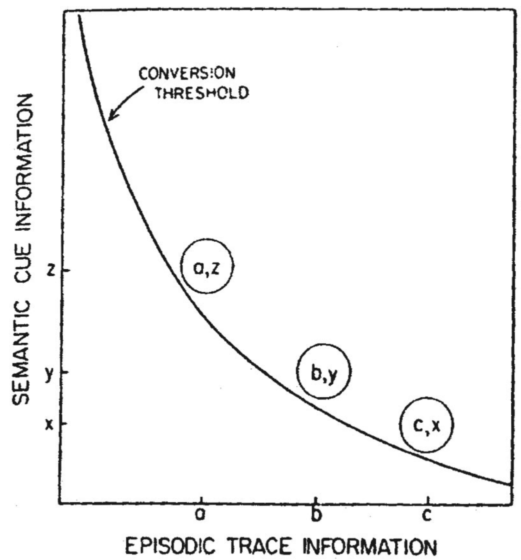


*Figure 1.* Tulving's (1985) representation of remembering and knowing. Test items differ in episodic and semantic information, and the relative strength of these two sources varies along a "conversion threshold." Events above the threshold are assigned an "old" response, those below it a "new" response. Events in the lower right region, like the point (*c, x*), are "remembered" whereas those in the upper left region, like the point (*a, z*) are "known." From "Memory and Consciousness," by E. Tulving, 1985, *Canadian Journal of Psychology, 26,* p. 7. Copyright 1985 by the Canadian Psychological Association. Reprinted with permission.

two-dimensional spaces shown in Figure 2. Old and new items differ, on the average, on both dimensions and can be represented as bivariate distributions; these are represented as circles of points having equal likelihood.

A decision variable corresponding to the sum of $x$ and $y$ can be represented in this space by a diagonal line with positive slope. If the sum is weighted so that the effective variable is $ax + y$, then the decision axis has slope $1/a$, as illustrated in Figure 3A. The space is divided into response regions by linear decision bounds; because these are perpendicular to the decision axis, they all have the negative slope $-a$. As one moves to the right along the decision axis, the value of a weighted sum of $ax + y$ increases.

In the one-dimensional model, $a = 1$ and there are three response regions (remember, know, and new), as shown in Figure 2A. All decisions depend on the same psychological variable, the weighted sum of $x$ and $y$. The distribution of such a sum can be found by projecting the bivariate distributions onto the decision axis or, equivalently, by convolution. The result is two univariate distributions differing in mean, and it is in these terms that the one-dimensional model is usually described. The shaded area is the region leading to an "old" response, which may be either "remember" or "know."

## *The Process-Pure Model*

We use the term *process-pure* to refer to remember–know models in which at least one of the responses depends on a single axis and thus "purely" depends on a single "process." The model described by Murdock (2006) is of this form: His Equations 3–6 are consistent with the representation in Figure 2B. A sufficiently high value of the $y$ variable (associative information, in his implementation) leads to a "remember" response; otherwise, the decision depends only on the $x$ variable (item information). The shaded region again corresponds to "old" responses. Similar models have been described by Reder et al. (2000) and by Yonelinas (2001), although in the latter case the $y$ information is considered to have a threshold rather than continuous character. Reder et al.'s (2000) SAC model assigns episodic strength to the $y$ variable and semantic strength to the $x$ variable (echoing Tulving, 1985), whereas Yonelinas (2001) prefers the terms *recollection* and *familiarity*.

Models in the process-pure category come the closest to capturing the popular model-free approach to interpreting remember–know data in which remember responses are assumed to directly tap a process of Type $y$, and know responses tap a process of Type $x$ (e.g., Gardiner & Richardson-Klavehn, 2000). Detection-theoretic process-pure models are more sophisticated in allowing for changes in response bias as the boundaries in Figure 2B shift (Macmillan, Rotello, & Verde, 2005). They resemble the model-free approach in assuming that both $x$- and $y$-type information contribute to the old–new decision, but only $y$ information is reflected in remember judgments.

## *STREAK*

The third model, STREAK (Rotello et al., 2004), assumes that all decisions rely on a combination of the two sources. STREAK agrees with the Wixted–Stretch model in basing the old–new decision on a sum of $x$ and $y$ information. It departs from that model in setting the remember–know decision bound orthogonal to the old–new bound so that remember–know judgments depend

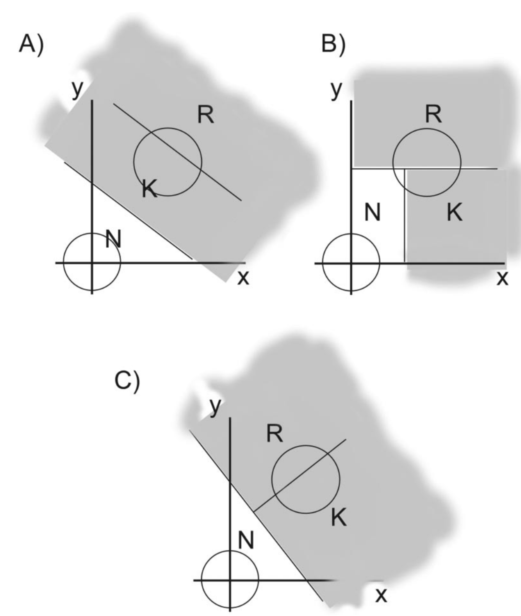


*Figure 2.* Modern quantitative models of remembering and knowing. Each model divides a two-dimensional space into regions for *remember* (R), *know* (K), and *new* (N) responses. The models are (A) the one-dimensional model (Wixted & Stretch, 2004), (B) the process-pure model (Murdock, 2006), and (C) STREAK (Rotello et al., 2004).

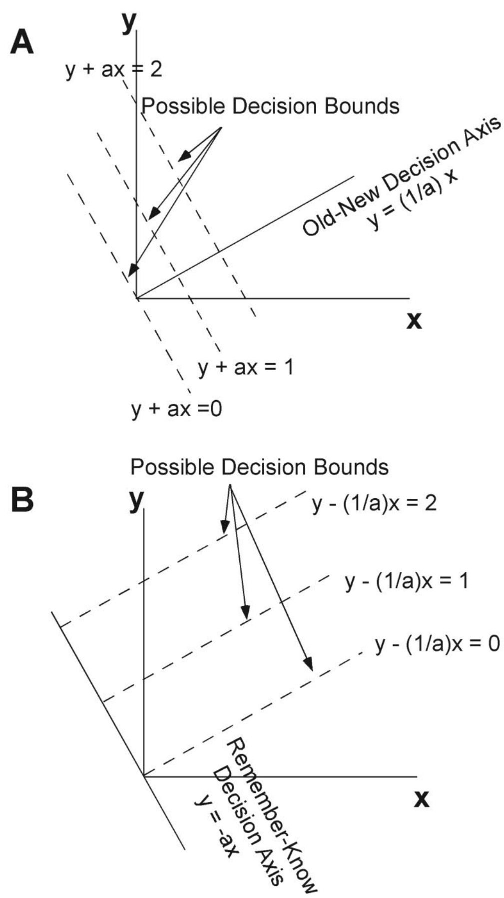


*Figure 3.* Decision variables in the *xy* plane that depend on a weighted sum or difference. (A) Linear decision bounds with negative slope imply a decision axis based on a weighted sum of $x$ and $y$. As one moves to the right along the decision axis, the value of $ax + y$ increases. Both the one-dimensional model ($a = 1$) and STREAK ($a = d_y/d_x$) use this decision rule. (B) Linear decision bounds with positive slope imply a decision axis based on a weighted difference between $y$ and $x$. As one moves upward along the decision axis, the value of $y - (1/a)x$ increases. STREAK ($a = d_y/d_x$) uses this rule.

on a difference between $y$ and $x$. As Figure 3B illustrates, a decision variable corresponding to such a difference can be represented in the decision space by a diagonal line with negative slope. If the difference is weighted so that the effective variable is $(y - [1/a]\,x)$ then the decision axis has slope $-a$. As one moves upward along this decision axis, the value of the weighted difference $(y - [1/a]\,x)$ increases. STREAK thus postulates an old–new decision based on a sum of $x$ and $y$ followed by a remember–know decision based on a difference; the resulting partition of the decision space is shown in Figure 2C, in which the "old" region is again shaded. As is evident when Figures 1 and 2 are compared, this model comes closest to capturing the original Tulving proposal.

Table 1 summarizes the decision variables assumed to underlie old–new and remember–know judgments in the three models. There is rough agreement about the former: old–new judgments depend on a combination of $x$ and $y$ information. In the one-dimensional model and STREAK, this combination occurs for each item and each judgment, whereas for the process-pure model it occurs across items. There is no agreement at all about remember–know judgments, which are interpreted by the three models to reflect a sum of $y$ and $x$, a difference between $y$ and $x$, or $y$ alone. One might expect that such starkly different ideas could be easily contrasted experimentally; we next consider Murdock's (2006) recommendations for accomplishing this task.

## Murdock's Solution: Connect the Model to Stimulus Characteristics Within a Process-Pure Model

Murdock (2006) developed a remember–know model inspired by the TODAM theory of memory (Murdock, 1982) and compared it favorably with STREAK and the one-dimensional model. He cited three major advantages of his analysis: (a) item and associative information are demonstrably independent and well-defined, so it is possible to identify empirical manipulations that correspond to changes along one strength axis but not the other; (b) TODAM is a process approach and is connected to a more general process model of memory; and (c) TODAM is presented more completely and is simpler than STREAK. The first point is the most substantive and requires the most discussion; in the end, however, we are largely unsympathetic to all three arguments.

### *Ability to Predict Results of Experimental Manipulations*

Murdock argued that stimulus manipulations lead to predictable changes in the parameters of his model. As evidence, he reanalyzed four classic experiments that display different types of interactions between remember and know responses. These experiments were selected by Dunn (2004) as fodder for his reanalysis, in which the one-dimensional model (the three-parameter equal-

Table 1
*Variables Assumed to Underlie Old–New and Remember–Know Decisions, According to Three Classes of Models*

| Model | Basis for old (rather than new) response | Basis for remember (rather than know) response |
|---|---|---|
| One dimensional | Sum of $x$ and $y$ exceeds a criterion | Sum of $x$ and $y$ exceeds a criterion |
| Process pure | Either $x$ or $y$ exceeds its own criterion | $y$ exceeds its criterion |
| STREAK | Sum of $x$ and $y$ exceeds a criterion | Difference between $y$ and $x$ exceeds a criterion |

*Note.* In specific models, $x$ is variously described as familiarity or as semantic, global, or item information; $y$ corresponds to recollection or episodic, specific, or associative information.

variance version) was successfully fit to the data. We now add STREAK to this mix.

For each of the four experiments, the parameter values estimated by TODAM are consistent with those of the one-dimensional model. Murdock concluded, however, that his model "gives a more detailed account of the data but of course at the cost of one more degree of freedom" (p. 652) This (accurate) caveat suggests that STREAK may be able to tell just as detailed a story for these data sets as does TODAM. Table 2 provides the parameter estimates of all three models for these data sets. Each model has a criterion parameter related to the "remember" response and a criterion parameter related to the "old" response. TODAM and STREAK have separate sensitivity parameters for $x$ and $y$ information, whereas the one-dimensional model has a single parameter for overall sensitivity. Let us consider the interpretations that Murdock's process-pure account, STREAK, and the one-dimensional model assign to these results.

Schacter, Verfaellie, and Anes (1997) compared normal controls and amnesic subjects. All models agree, unsurprisingly, that normal subjects have higher sensitivity. STREAK and TODAM agree that sensitivity is greater for the $y$ dimension than for the $x$ dimension for normal subjects but not for amnesic subjects. The interesting difference is in the pattern of criteria: TODAM and the one-dimensional model conclude that both criteria are more conservative for normal controls, although the effect is smaller for the remember criterion. STREAK agrees that the old–new criterion is higher but finds that the remember–know criterion is lower. Intuitively, this result arises because in STREAK (but not its competitors) a change in the old–new criterion perforce produces a change in remember responses. A finding consistent with the STREAK interpretation, discussed later, is that the remember rate does change systematically when old–new response bias is manipulated (Rotello, Macmillan, Hicks, & Hautus, in press).

Gregg and Gardiner (1994) contrasted auditory and visual presentations; remember responses were unaffected, but visual items more often received a know response. For these data, all models agree on the interpretation: Sensitivities are greater for visual presentation, and the two criteria shift in opposite directions as a function of modality.

Gardiner and Java (1990) compared memory for words and nonwords. TODAM concludes that the largest change is in sensitivity to the $x$ dimension, with other parameters changing by smaller amounts. STREAK and the one-dimensional model find the most substantial difference to be a more conservative value of the remember–know criterion for nonwords. Dunn (2004) chose this data set for analysis because remember and know responses changed in opposite direction between conditions, and the conclusions reached by STREAK and the one-dimensional model are less surprising (although not necessarily more correct) than those reached by TODAM.

In the fourth study, Gardiner, Kaminska, Dixon, and Java (1996) contrasted one- and four-trial conditions. The models agree that four study trials lead to higher sensitivity and at least nominally higher criteria. TODAM and the one-dimensional model infer that the larger criterion shift is for the remember criterion, whereas STREAK finds that essentially all of the bias difference is in the old–new criterion. As with the Schacter et al. (1997) data, this outcome is a result of STREAK's unique claim that remember responses are affected by a change in the old–new criterion.

Table 2
*Values of Parameters From TODAM, STREAK, and the One-Dimensional (1D) Model for Five Data Sets*

| | | Criterion | | | | | | Sensitivity | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | Remember | | | Old | | | Type $y$ | | Type $x$ | | Overall |
| Study | Condition | *TODAM* ($\hat{a}$) | *STREAK* ($-C_r$)[a] | 1D ($r$) | *TODAM* ($\hat{b}$) | *STREAK* ($C_o$) | 1D ($k$) | *TODAM* ($\mu[f_o]$) | *STREAK* ($d_y$) | *TODAM* ($\mu$ $[g_o]$) | *STREAK* ($d_x$) | 1D ($d'$) |
| Schacter et al. (1997) | Amnesic | 1.55 | 0.41 | 1.62 | 0.91 | 0.59 | 0.75 | 0.51 | 0.35 | 0.50 | 0.28 | 0.60 |
| | Normal Control | 1.88 | −0.30 | 1.98 | 1.39 | 0.98 | 1.25 | 1.75 | 1.51 | 1.41 | 0.94 | 1.85 |
| Gregg & Gardiner (1994) | Auditory | 1.64 | 0.61 | 2.12 | 1.31 | 0.86 | 1.19 | 0.36 | 0.31 | 0.79 | 0.54 | 0.86 |
| | Visual | 1.88 | 0.94 | 2.43 | 0.89 | 0.65 | 0.88 | 0.65 | 0.66 | 1.10 | 0.72 | 1.21 |
| Gardiner & Java (1990) | Word | 1.75 | −0.35 | 1.55 | 1.20 | 0.83 | 1.07 | 1.17 | 1.02 | 0.44 | 0.36 | 0.95 |
| | Nonword | 1.88 | 0.29 | 1.89 | 1.16 | 0.83 | 1.04 | 1.00 | 0.76 | 0.83 | 0.47 | 1.01 |
| Gardiner et al. (1996) | 1 Trial | 1.23 | 0.06 | 1.45 | 1.01 | 0.54 | 0.70 | 0.65 | 0.56 | 0.84 | 0.52 | 0.87 |
| | 4 Trials | 2.05 | 0.11 | 2.56 | 1.33 | 0.98 | 1.30 | 1.74 | 1.64 | 1.84 | 1.20 | 2.25 |
| Hockley & Consoli (1999) | Items | 1.24 | −0.45 | 1.28 | 0.75 | 0.47 | 0.64 | 0.94 | 1.27 | 0.51 | 0.48 | 1.12 |
| | Pairs | 1.51 | −0.62 | 1.23 | 0.69 | 0.43 | 0.62 | 1.53 | 1.47 | 0.46 | 0.50 | 1.19 |

[a] Negative values of $C_r$ are given to scale in the same direction as $\hat{a}$ and $r$.

We provide these plausibility arguments to show that STREAK and TODAM cannot be distinguished by this approach, but the conclusions that can be reached by this avenue are weak. All the stories told by these models are to some degree plausible. A potentially more revealing study would be one in which the putative dimensions of Murdock' model, item and associative information, were experimentally manipulated. We have added a fifth study to the testbed that satisfies this requirement, and if the TODAM model is correct it should be particularly successful in fitting the data. Hockley and Consoli (1999) tested subjects on item and pair (associative) recognition at three test delays in each of two experiments. We averaged their data over delay and experiment then fit the results with the competing models (see Table 2). TODAM infers, appropriately, that $y$-type (associative) sensitivity is increased for pairs compared with items and that $x$-type (item) sensitivity is essentially unchanged. In addition, the remember criterion is more conservative for pairs. However, STREAK reaches exactly the same conclusions about sensitivity (the two models find small changes in the remember criterion, in opposite directions). The one-dimensional model, paying the price for having fewer parameters, concludes that there are no substantial differences between items and pairs. Even when the experimental variables of the TODAM model are directly manipulated, that model provides neither special insight nor a more clear-cut result.

Not only does this plausibility exercise fail to distinguish among the decision models, it also fails to demonstrate the identity of dimensions that form the decision space. Consider the Gardiner et al. (1996) analyses, for which both TODAM and STREAK find greater $x$- and $y$-sensitivity in the four-trial condition. What would constitute evidence that it is item and associative strength—rather than global and specific strength, semantic or episodic memory, familiarity or recollection—that accounts for this result? All are equally plausible until some independent measure of these terms is provided. Many experiments have produced changes in remembering in the absence of obvious changes in associative information across conditions. A more neutral view of how the decision space is to be defined avoids the temptation to believe something has been explained that has merely been named.

## *What Does TODAM Have to Do With It?*

This observation brings us to Murdock's second point, that an advantage of his remember–know model is its relation to TODAM, a process model. In TODAM, memory storage is described in terms of a random vector on which encoding and retrieval processes operate. Specifying a process in this way is an important accomplishment, but very little of the mechanics of TODAM are brought to bear on the remember–know problem and the relation between the new model and its parent is rather arm's-length.[2] Murdock's Equations 3–6 defined a decision rule, and his allusions to TODAM did not turn that rule into a process. We have already seen that many alternative axis labels are consistent with that rule; if the process-pure model proves successful, no benefit will accrue to TODAM or to the labeling of the decision axes preferred by Murdock (2006).

Three factors serve to undo the greater purchase provided by tying the decision model to stimulus factors. First, those stimulus factors offer no help in the many experiments (like Studies 1–4 in Table 2) in which variables other than item and associative strengths are manipulated. Second, in ideal experiments like that of Hockley and Consoli (1999), the TODAM model demonstrated no advantage. Third, and perhaps most seriously, nothing in the TODAM memory model (Murdock, 1982) forces a process-pure decision model for the remember–know task. Either of the other two partitions of the decision space shown in Figure 2 would be equally compatible with the Item × Associative stimulus space.

The one advantage that Murdock's preferred decision axes do have is their demonstrable independence. The converse of this statement, however—that a disadvantage of STREAK is questionable independence—is not correct. STREAK's failure to make a commitment about the nature of the axes is a plus here: The terms *global* and *specific* do not refer to larger models of memory and do not imply (as Murdock has suggested) a part–whole relation or any other kind of dependence. In our interpretation, STREAK proposes that two independent variables underlie recognition in the remember–know paradigm, but the names of the axes are effectively left to the reader.

## *Completeness and Simplicity of Models*

We now turn to two specific criticisms of STREAK raised by Murdock: that it is incomplete and that it is unduly complex. Neither criticism is justified.

*Model completeness.* Murdock (2006) asserted that "integral equations are not provided" for STREAK (p. 649). He pointed to his own Equations 3–6 as examples of such equations. Those equations express the remember and know hit and false-alarm rates as integrals of normal distributions according to TODAM. Our Equations B1 to B4 (Rotello et al., 2004, p. 616) express the remember and overall (Remember + Know) hit and false-alarm rates as integrals of normal distributions according to STREAK. The only difference is in our notation for writing normal integrals. He wrote an explicit integral of a normal density, such as

$$P(\text{“}remember\text{”}|New) = \int_{u=a}^{\infty} f_N(u)du, \tag{1}$$

where $f_N$ indicates a unit normal density. However, "it is traditional to denote the cumulative distribution function of a standard normal distribution by $\Phi(x)$" (Ross, 1988, p. 165), and that is the notation we used. So, for example, our Equation B1 was

$$P(\text{“}old\text{”}|New) = \Phi(-C_O/s). \tag{2}$$

But this is by definition equivalent to

$$P(\text{“}old\text{”}|New) = \int_{-\infty}^{-C_O/s} f_N(u)du, \tag{3}$$

which is in exactly the same form as Murdock's equations. (Indeed, Murdock himself used $\Phi(x)$ in his Equations 7–10.) Although other aspects of our disagreement might be viewed as

[2] In contrast, Reder et al., 2000, provide an excellent example of how the processes that underlie R–K judgments can be modeled.

subjective, his assertion that we failed to include integral equations is an error of fact.

*Model simplicity.* Murdock (2006) remarked that his model is simpler than STREAK in allowing for direct calculation rather than more complex computational methods. He found the grid-search parameter estimation procedure used by Rotello et al. (2004) inferior, and claimed that "one must rely on simulations to check both one's understanding of the model and the accuracy of the model derivations" (p. 8).

*Well.* We do agree that better estimation methods are desirable and are now able to provide equations to fit STREAK to (nonrating) remember–know results (see Appendix).[3] The availability of these equations renders STREAK as calculationally "simple" as any other model. As for the role of simulations, many techniques have heuristic value as aids to understanding, but we believe that the acid test of the accuracy of analytic derivations is to do the math.

In a somewhat different sense, Murdock's model may be too simple. A more general version of the process-pure model in which old and new distributions are allowed to have unequal variance may be necessary to account for normalized receiver operating characteristic (*z*-ROC) slope and other aspects of recognition data. The Appendix presents equations similar to Murdock's Equations 3–6 that describe this case.

## Distinguishing Among the Models

Murdock (2006) laid out a new model rather than improving on old ones and thus presented the challenge of deciding among a growing set of alternative models. We argue that relating the models to stimulus variables has not yet been a decisive tool, but we do not believe the problem is insurmountable. We suggest three helpful strategies for choosing among them.

1. Extend the models to rating designs.

As Murdock pointed out, deciding among the models based on their ability to account for typical remember–know data (like those summarized in Table 2) is not a purely statistical enterprise: In their basic form, STREAK and TODAM are saturated and can fit data perfectly. One way to "desaturate" the models is by the use of response ratings to generate ROC curves. Ratings may be applied to any binary response contrast, and we have used them to expand both the old–new and the remember–know response (Rotello et al., 2004; Rotello, Macmillan, Reeder, & Wong, 2005; Rotello et al., in press). This approach increases the degrees of freedom in the data so that the models may be compared quantitatively, with corresponding measures of goodness of fit (Rotello et al., in press).

Rating experiments are usually modeled by interpreting differences in confidence as differences in decision boundary locations, and Rotello et al. (2004, in press) are explicit about these locations. Murdock did not tell us how his model would be extended to ratings designs. The most common remember–know design, in which an old–new rating is followed by a binary remember–know response, does not seem natural for the process-pure decision bounds, as Murdock acknowledged. One aspect of rating data that any model must account for is the slope of the old–new recognition ROC (on $z$ coordinates), which is generally found to be $< 1$. Murdock (2006) speculated that the addition of encoding variability might allow his model to predict shallower slopes for ratings data, but whether this can be accomplished without interfering with other predictions is, as he notes, unknown. Another approach is to copy STREAK in simply assuming that the old distribution has a greater variance than the new distribution (as is done in the Appendix for the nonrating paradigm).

2. Examine predicted response patterns.

Each model makes a specific prediction about one aspect of the data. STREAK predicts a *response-ratio invariant*: When old–new bias is manipulated across conditions, the proportion of "remember" judgments should be a constant proportion of the "old" responses. Support for this prediction has been presented in Rotello et al. (in press) and in six of seven past experiments summarized in that article. The prediction arises from the orthogonal arrangement of decision bounds in STREAK; recall that in fitting the data in Table 2, this feature led to interpretations that were different from those of other models. The similar arrangement of orthogonal decision bounds in the process-pure model leads to a related prediction by that model: when remember–know bias changes across conditions the proportion of "know" judgments should be a constant proportion of the "not remember" responses. Few experiments have actually manipulated subjects' bias to say "remember," but Rotello et al. (2005, Experiment 2) did observe this constancy, finding $P(\text{know}|\text{not remember}) = .29$, with conservative remembering, and $P(\text{know}|\text{not remember}) = .30$, with neutral remembering.

The one-dimensional model's signature prediction compares the two-point $z$-ROC that can be obtained from remember–know judgments with the multipoint $z$-ROC obtained from ratings. Because all judgments depend on the same decision axis, the slopes of these two functions should be equal. Rotello et al. (2004) found that across many studies, two-point slopes averaged about 1.0, reliably more than the 0.8 typically found for old–new rating tasks. To address this problem, Wixted and Stretch (2004) proposed a variant of the one-dimensional model in which the remember–know criterion was variable. The model allows two-point and rating slopes to differ, and leads to the weaker prediction that slope estimates are highly correlated across individuals and conditions. In an experimental test, however, this version of the model was less successful than its fixed-criterion cousin (Rotello et al., in press). Malmberg and Xu (in press) have suggested that between-subjects comparisons of slopes are inappropriate, but Rotello et al. (in press) found the usual discrepancy in a within-subject test. Two-point slopes remain problematic for the one-dimensional model.

3. Match models to variants in experimental designs.

The most common remember–know task requires an old–new decision followed (in the case of an "old" response) by a remember–know judgment; such a task corresponds naturally to the STREAK decision bounds (Figure 2C) and the one-dimensional bounds (Figure 2A). An alternative version of the paradigm is to ask subjects to first decide whether a test probe is "remembered" and, if not, whether it is a new item or an old item

[3] We thank John Dunn for providing this derivation.

that is "known." This sequence of decisions maps nicely onto the process-pure decision bounds (Figure 2B) and the one-dimensional model.

A potentially instructive research strategy is to compare the structure of data for different remember–know paradigms. That there are some differences is known: We compared experiments in the Rotello et al. (2004) database that used the old–new followed by remember–know paradigm (ON–RK) with those having an evenhanded trinary response set (RKN). On average, RKN experiments produced more hits, shallower two-point slopes, and lower estimated criteria than ON–RK experiments. It may turn out that the decision partitions proposed by the process-pure model and by STREAK can each be used in particular experimental designs.

## Conclusion

The three types of models that account for remember–know judgments—one-dimensional, process-pure, and STREAK—each have clear advantages. For example, the one-dimensional model provides a good qualitative description of many aspects of remember–know data (Dunn, 2004; Wixted & Stretch, 2004) and a good quantitative fit to ROC data (Rotello et al., in press). STREAK accounts for the two-point ROC slopes that are problematic for that model, accurately predicts a response-ratio invariant under bias changes, and also provides a good quantitative fit to remember–know rating ROCs (Rotello et al., in press). TODAM ties remember–know judgments to other types of memory tasks and provides clearly manipulable labels for the strength axes. No single criterion can be used to evaluate models, and it may be that no current model can account for all findings to date. Perhaps the true model will contain the present ones as special cases.

What is clear is the great advantage quantitative modeling has over the model-free approach that dominated the early use of the remember–know paradigm. It is interesting to compare the use of this paradigm with the use of ratings instruments, both of which ask subjects to report aspects of their subjective experience. Rating data are now routinely analyzed by using signal detection theory and in that garb have proved a powerful research tool. The era of naive interpretation of remember and know responses may be ending, but model-based understanding of such reports, and other similar ones, has theoretical promise.

## References

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

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

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.

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.

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.

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

Macmillan, N. A., Rotello, C. M., & Verde, M. F. (2005). On the importance of models in interpreting remember–know experiments: Comments on Gardiner et al.'s (2002) meta-analysis. *Memory, 13,* 607–621.

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

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

Murdock, B. B. (2006). Decision-making models of remember–know judgments: Comment on Rotello, Macmillan, and Reeder (2004). *Psychological Review, 113,* 648–656.

Reder, L. M., Nhouyvanisvong, A., Schenn, 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.

Ross, S. (1988). *A first course in probability (3rd ed.)*. New York: Macmillan.

Rotello, C. M., Macmillan, N. A., Hicks, J. L., & Hautus, M. J. (in press). Interpreting the effects of response bias on remember–know judgments using signal-detection and threshold models. *Memory & Cognition.*

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. M., Macmillan, N. A., Reeder, J. A., & Wong, M. (2005). The remember response: Subject to bias, graded, and not a process-pure indicator of recollection. *Psychonomic Bulletin & Review, 12,* 865–873.

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

Tulving, E. (1985). Memory and consciousness. *Canadian Journal of 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. P. (2001). Consciousness, control, and confidence: The 3 Cs of recognition memory. *Journal of Experimental Psychology: General, 130,* 361–379.

(*Appendix follows*)

# Appendix

## Equations for Finding Parameters of STREAK and of the Unequal-Variance Process-Pure Model From Response Proportions

### STREAK

Rotello et al. (2004) provided equations that express response proportions in terms of model parameters; this appendix provides the inverse result. For the parameters, we adopt the notation of Rotello et al.: Global and specific sensitivity are denoted by $d_x$ and $d_y$, the old–new and remember–know criteria by $C_o$ and $C_r$. The data may be summarized by the hit and false-alarm rates (H and F) and the remember hit and false-alarm rates (*RH* and *RF*).

To follow the proof, it is helpful to consult Figure 4 in Rotello et al. (2004), which is an expanded version of Figure 2C in the present article. Some helpful notation is as follows: The distance between the means of the New and Old distributions is, by the Pythagorean theorem, $\sqrt{d_x^2 + d_y^2}$, which we denote as $D$. The distance $D$ can also be partitioned into two components in another way, as the Pythagorean sum of distance along the old–new decision axis ($d_o$) and along the remember–know decision axis ($d_r$). Thus it is also true that

$$D = \sqrt{d_o^2 + d_r^2} = \sqrt{d_x^2 + d_y^2}\,. \tag{A1}$$

It can be shown by geometry that

$$d_o = \frac{2d_x d_y}{\sqrt{d_x^2 + d_y^2}} \tag{A2}$$

and

$$d_r = \frac{d_x^2 - d_y^2}{\sqrt{d_x^2 + d_y^2}}\,. \tag{A3}$$

Comparison with Appendix B of Rotello et al. (2004) shows that

$$C_o = (-s)z(F)$$

$$C_r = z\left(\frac{RH}{H}\right)$$

$$d_o = z(H) + C_o$$

$$d_r = C_r - (s)z\left(\frac{RF}{F}\right), \tag{A4}$$

where $s$ is the standard deviation of the new distribution, and $z$ is the inverse of the normal distribution function. Equations A1 and A2 can be solved for $d_x$ and $d_y$, with the result

$$d_x = \sqrt{\frac{1}{2}(D^2 - d_r D)}$$

$$d_y = sign(d_o)\sqrt{\frac{1}{2}(D^2 + d_r D)}\,. \tag{A5}$$

Equations A1, A4, and A5 can be used to find the four model parameters. Although the algebra is more complex than for the process-pure model, the calculation is easily done on a spreadsheet and no simulation or parameter search is needed.

### An Unequal-Variance Process-Pure Model

Murdock's (2006) model is summarized in his Equations 3–6, which give expressions for the remember hit and false-alarm rate (here denoted *RH* and *RF*) and the know hit and false-alarm rate (*KH* and *KF*). The model is easily extended to allow for unequal variance of the Old and New distributions. As with STREAK, we set the standard deviation of the New distribution to $s$ and that of the Old distribution to 1. Expressions corresponding to Murdock's Equations 3–6 are

$$RF = \Phi\left(-\frac{a}{s}\right)$$

$$RH = \Phi(d_y - a)$$

$$KF = (1 - RF)\Phi\left(-\frac{\mathrm{b}}{\mathrm{s}}\right)$$

$$KH = (1 - RH)\Phi(d_x - b)\,, \tag{A6}$$

where $\Phi$ is the normal distribution function, $a$ is the remember criterion and $b$ is the know criterion. Equation A6 can be readily solved for the parameters:

$$a = (-s)z(RF)$$

$$b = (-s)z\left(\frac{KF}{1 - RF}\right)$$

$$d_y = z(RH) + a$$

$$d_x = z\left(\frac{KH}{1 - RH}\right) + b. \tag{A7}$$

We fit the unequal-variance version of the process-pure model to the data in Table 2 and found that, in all cases, the criteria $a$ and $b$ and the sensitivities $d_x$ and $d_y$ were lower than if equal variance were assumed. The discrepancy is reduced if these parameters are scaled in units of the New rather than the Old distribution. The qualitative pattern of results is similar for both variants.

Received November 16, 2005
Revision received January 9, 2006
Accepted January 10, 2006 ■

---

# Postscript

Neil A. Macmillan and Caren M. Rotello
*University of Massachusetts at Amherst*

We are in partial (but only partial) agreement with all the points in Murdock's (2006) postscript. To take his comments in order:

1. Certainly it is important to connect remember–know experiments with other work on recognition memory, although we have argued that Murdock's approach is not uniquely successful in doing so.

2. The decision axes in STREAK are indeed rotated from the strength axes, but the model as a whole is not simply