The influence of averaging and noisy decision strategies on the recognition memory ROC

KENNETH J. MALMBERG and JING XU
Iowa State University, *Ames, Iowa*

Many single- and dual-process models of recognition memory predict that the ratings and remember–know receiver operating characteristics (ROCs) are the same, but Rotello, Macmillan, and Reeder (2004) reported that the slopes of the remember–know and ratings z-transformed ROCs (zROCs) are different. The authors show that averaging introduces nonlinearities to the form of the zROC and that ratings and remember–know zROCs are indistinguishable when constructed in a conventional manner. The authors show, further, that some nonoptimal decision strategies have a distinctive, nonlinear effect on the form of the single-process continuous-state zROC. The conclusion is that many factors having nothing to do with the nature of recognition memory can affect the shape of zROCs, and that therefore, the shape of the zROC does not, alone, characterize different memory models.

At issue is the nature of the information used to recognize whether a stimulus was encountered in a specific context. *Single-process* continuous-state models assume that recognition is based on a continuous random variable that is often conceptualized as familiarity (see, e.g., Egan, 1958). The more similar the stimulus is to memory traces associated with a specified context, the more familiar the stimulus seems and the more likely it is that one will positively endorse it. *Dual-process* models assume that recognition is sometimes based on stimulus familiarity and other times based on recalling the details of an event (see, e.g., Atkinson & Juola, 1974; Jacoby, 1991; Malmberg, Holden, & Shiffrin, 2004; Mandler, 1980; Reder et al., 2000; Rotello, Macmillan, & Reeder, 2004; Yonelinas, 1994). If the details retrieved from memory correspond to the specified context, the stimulus is endorsed. If details are not retrieved, recognition is based on stimulus familiarity.

The remember–know task (RK task; Gardiner, 1988; Tulving, 1983) has been purported to distinguish between the single- and dual-process models (Yonelinas, 2002). Following a positive yes–no recognition decision, subjects introspectively decide whether their "yes" response was based on recollecting episodic details or on a feeling of knowing that they had studied the item. We refer to the probability of responding "yes" to a studied item

(or target) as the hit rate (HR) and the probability of responding "yes" to an unstudied item (or foil) as the false alarm rate (FAR). Likewise, we refer to the probability of responding "remember" to a target as the remember HR and the probability of responding "remember" to a foil as the remember FAR. Accordingly, there typically exist two HR–FAR pairs for each condition in a remember–know experiment, and the function relating these data is referred to as the remember–know receiver operating characteristic (RK ROC).

Many single-process models of the RK task are extensions of signal detection theory (Donaldson, 1996; Green & Swets, 1966; Hirshman & Master, 1997; Malmberg, Zeelenberg, & Shiffrin, 2004). Recognition is above chance when the mean of the target familiarity distribution is greater than the mean of the foil familiarity distribution, and the RK task is modeled by comparing a sample from the appropriate distribution to two criteria. If the familiarity of the stimulus exceeds a relatively lenient yes–no criterion, the response is "yes." If the familiarity of the stimulus subsequently exceeds a stricter RK criterion, the response is "remember"; the response is "know," otherwise. Within the dual-process framework, variables affecting remember and know rates, respectively, reflect changes in the contribution of recollection and familiarity to recognition.

The ratings task has also been purported to distinguish between single- and dual-process models (Yonelinas, 2002). In a ratings task, subjects determine their confidence that they had studied a test item. For each level of confidence there is an HR–FAR pair, and the cumulative function relating HR–FAR pairs as a function of confidence is the ratings ROC. For many continuous-state models, the ratings ROC is nonlinear, and when it is z-transformed, it produces a linear zROC with a slope less than 1.0. Single-process continuous-state models predict that the ratings and RK ROCs are identical, because the only difference

between the ratings and the RK models is the number of criteria employed to perform the task.

According to some dual-process models (Yonelinas, 1994), the highest confidence "old" response is given only when recollection is the basis for the old–new decision. Otherwise, confidence ratings are based on the test item's familiarity. On this assumption, the HR–FAR pair that corresponds to the strictest criterion in a ratings experiment is the same as the probability of responding "remember" to targets and foils in an RK experiment. Thus, these dual-process models predict identical asymmetrical ROCs and slightly U-shaped zROCs for RK and ratings tasks. In contrast, Rotello et al. (2004) concluded that the ratings ROC is more asymmetrical than the RK ROC, on the basis of a meta-analysis of more than 300 RK and more than 100 ratings conditions from a variety of experiments. That is, the average slope of the RK zROC was greater than the average slope of the ratings zROC.

Different RK and ratings zROCs are an empirical challenge for many recognition models, but before we reject them, it is important to note that this finding does not align with several other recent findings. For instance, Wixted and Stretch (2004) reviewed several experiments and found that the slopes of the ratings and RK zROCs were similar. Dunn (2004) and Malmberg, Zeelenberg, and Shiffrin (2004) showed that a wide range of RK findings are handled by single-process models, and Reder et al. (2000) showed that a dual-process model fits RK ROCs. Thus, single- and dual-process models describe the RK ROC, but it is unclear why the ratings and RK zROCs at least sometimes have different slopes.

Here, we describe and empirically test a simple account. It is critical that Rotello et al.'s (2004) meta-analysis was based on group mean data, whereas the convention is to derive slope estimates from individual zROCs. Rotello et al.'s reliance on group mean zROCs was certainly due to the heroic nature of undertaking a review of such a large number of studies. However, there are well-documented dangers associated with relying on the performance of the average individual to infer individual performance (Estes, 1956, 2002). For instance, Macmillan and Kaplan (1985) wrote that in situations in which individual performance is unknown, careful application of detection theory is required in order to characterize individual performance based on averaged data. It is important to squarely address the averaging issue, because if the RK ROC is truly different from the ratings ROC, many otherwise successful models are disconfirmed.

According to the averaging account, at least some of the zROCs in the meta-analysis were distorted into an inverted U-shaped form by averaging, and their RK zROCs tended to have a steeper slope than their ratings zROCs, because their RK zROCs were confined to a stricter portion of an inverted U-shaped zROC. We further note that nonlinear zROCs are often reported, and we describe, within the framework of a single-process model, several other ways that the form of the zROC can be distorted by decision strategies and not by memory processes.

On the Form of the Recognition zROC

Here we show that the averaging method can drastically distort the slope of the RK zROC. Consider that the relationship between probabilities and z scores is not linear, and that small changes in probabilities near 0 and 1 produce very large changes in z-transformed probabilities, but equivalent changes in probabilities near .50 produce relatively small changes when they are z-transformed. Thus, moving equal distances at different points on the probability scale does not necessarily produce equivalent changes in z scores. We illustrate this with an example.

Assume that one observed two subjects (S_1 and S_2) in an RK experiment. For the first point on the RK ROC, the two observations for each subject are the probabilities of responding "remember" when targets ($S_1 = .50$ and $S_2 = .70$) and foils are tested ($S_1 = .001$ and $S_2 = .20$). Again, we will respectively refer to these probabilities as the remember HR and the remember FAR. Averaging over S_1 and S_2, the mean remember HR is .60 and the mean remember FAR is .1005.

The stricter point on the RK zROC is determined by z-transforming the remember HR and FAR. This may be done in two ways, but the two ways clearly produce different results. The mean of the z-transformed HRs and FARs for S_1 and S_2 is the conventional method. The averaging method uses the z scores of the mean remember HR and FAR for S_1 and S_2. The mean z score of the two remember FARs is −1.966, whereas the z score of the mean FAR (.1005) is −1.278. The mean z score of the remember HRs is .261, but the z score of the mean remember HR (.60) is .253. Thus, the averaging method creates a shift in the remember point on the RK zROC from (−1.966, .261) to (−1.278, .253), where the first element of each z score pair corresponds to the remember FAR and the second to the remember HR.

Figure 1 illustrates the averaging shift. The means of the z scores are plotted with white circles in Figure 1, and the z scores of the mean remember HR and FAR are plotted with black circles. The arrows that are labeled P(remember) indicate that the black circle is shifted far to the right of the white circle, but there is only a small downward shift in the z scores of the remember HRs. Thus, averaging over the remember HRs and FARs before z-transforming them severely distorted the location of the strictest point on the RK zROC. This distortion can be thought of as a type of external noise introduced to the zROC. Whether it is internally or externally generated, noise always reduces performance; hence, the extreme ends of the averaged zROC will approach chance performance to a greater extent than will intermediate points on the zROC (cf. Macmillan & Kaplan, 1985).

The second, more lenient point on the RK zROC is simply derived from the HRs ($S_1 = .70$, $S_2 = .90$) and the FARs ($S_1 = .20$, $S_2 = .40$), whose means are .80 and .30, in this example, respectively. The HRs and FARs fall more centrally on the probability scale, and hence, the averaging method produces only a small shift, from (−.547, .853) to (−.524, .842). This is illustrated in Figure 1 with
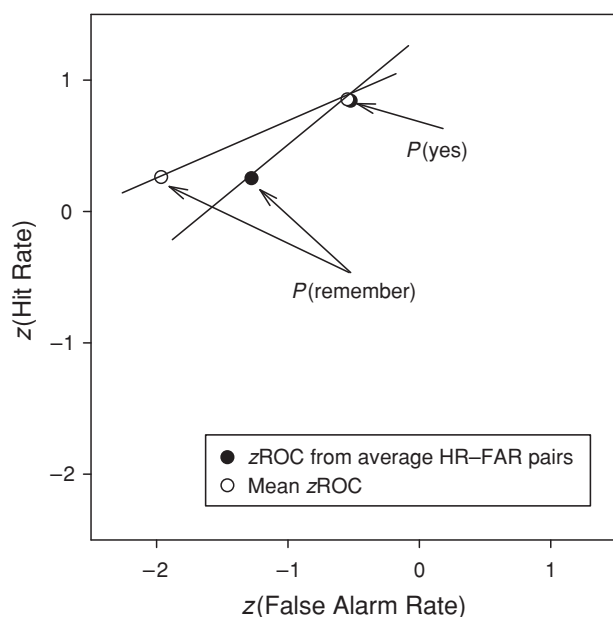
Figure 1. The effect of averaging on *z*ROC RK slope estimates. Note that the slope of the *z*ROC generated from the average HR–FAR pairs is greater than the slope of the mean *z*ROC.

the largely overlapping white and black circles that are labeled *P*(yes).

The different methods for creating *z*ROCs can lead to RK *z*ROCs with different slopes. One can see this in Figure 1 by linearly connecting the two points on each *z*ROC. The slope of the *z*ROC constructed using the averaging method is steeper than the *z*ROC based on average *z* scores. Because Rotello et al. (2004) used the averaging method to generate their *z*ROCs, they almost certainly overestimated the slope of the RK *z*ROC.

For the ratings task, the averaging method produces an inverted U-shaped *z*ROC, because averaging causes greater distortions at the extremities of the *z*ROC. Although averaging distorts the form of the ratings *z*ROC, it has a relatively small effect on the slope of the ratings *z*ROC when compared with the effect of averaging on the slope of the RK *z*ROC. To see why, note that the primary differences between the RK and ratings ROCs are the number of points and the position of the points on the inverted U-shaped *z*ROC. Whereas RK *z*ROCs are typically formed from two points that are positioned on the conservative portion of *z*ROC space, there are more points on the ratings *z*ROC, and they are more evenly distributed in *z*ROC space. The wider distribution of points on the ratings *z*ROC protects its slope, to some degree, from averaging.

The Present Experiment

The preceding analysis shows that the averaging method overestimates the slope of the RK *z*ROC, and this might explain why the average slope of the RK *z*ROC in Rotello et al.'s (2004) study was near 1.0. Furthermore, the ad-

dition of more lenient points to the *z*ROC will attenuate the overestimation, meaning that averaging is therefore a plausible explanation for the reported difference in the slope estimates of the RK and ratings *z*ROCs. The critical empirical question is whether analyzing the mean of the individual *z*ROC slopes, as opposed to the *z*ROC slope generated from the mean HRs and FARs, affects the slope estimates in practice.

METHOD

Subjects

Thirty-nine students enrolled in undergraduate psychology courses at Iowa State University participated in exchange for extra course credit.

Design and Materials

Each subject studied four 30-item lists consisting of nouns sampled from the Francis and Kučera (1982) norms. The range of normative word frequencies varied from 1 to over 1,000 per million. Assignment of words to lists and the target–foil condition was determined randomly for each subject. Each word was studied for 1.5 sec. Following each study list, a 20-sec distractor task was performed, which consisted of mentally adding digits. Following the distractor task were 30 target and 30 foil test trials in a random order that was determined anew for each subject.

Procedure

The subjects were told that sometimes an item could be recognized as having been studied because they could recall some of the details of the study event, and that sometimes they could recognize a word because it seemed relatively familiar to them, even though they could not recall some of the details of having studied the word. The former basis for recognition was likened to meeting a person on the street and recognizing that person from a recent party or class one had attended. The latter basis for recognition was likened to meeting a person on the street and not remembering where one had met this person, but nevertheless knowing that one had met the person before.

For each test trial, the subjects rated, on a scale of 1 to 4, how confident they were that they had studied the test word. Responses of "1" and "2," respectively, indicated high and moderate confidence that they had studied the word, and responses of "4" and "3," respectively, indicated high and moderate confidence that they had not studied the word. After a positive ratings judgment, subjects gave an RK judgment by responding "r," if they based their initial response on remembering that they had studied the item, or "f," if they made their response because the word seemed relatively familiar to them. Following a negative ratings judgment, subjects responded "r," if they based their initial response on remembering that they had not studied the item (i.e., they last encountered the word before the experiment), or "f," if they made their response because the word seemed relatively unfamiliar to them.

RESULTS AND DISCUSSION

The results of one subject were eliminated from the analyses because an incomplete ROC was produced. The means of the HRs and FARs for the ratings and RK ROCs are plotted in Figure 2. None of the probabilities used to construct the RK and ratings ROCs deviate by more than .012, indicating that the ratings and RK ROCs are statistically the same ($p > .05$).

The memory models with which we are working are the single-process retrieving effectively from memory
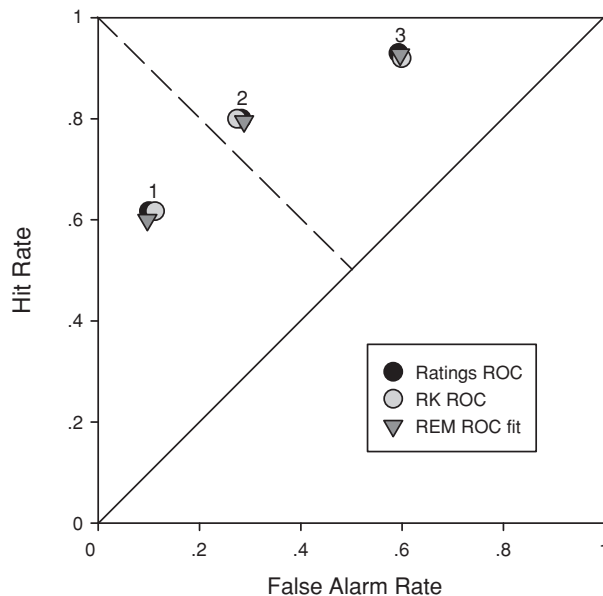
Figure 2. Results of the experiment and a fit of the single-process REM model. The ROCs are the cumulative probabilities of responding at different levels of confidence for the ratings ROC or remembering versus knowing for the remember–know (RK) ROC. The HR–FAR labeled 1 corresponds to the probability of using the highest confidence "old" response for the ratings ROC and the probability of an "old remember" judgment on RK ROC. The HR–FAR labeled 2 corresponds to the probability of using the highest confidence "old" response or the lowest confidence "old" for the ratings ROC and the probability of an "old remember" judgment or an "old know" judgment on RK ROC. The HR–FAR labeled 3 corresponds to the probability of using the highest confidence "old" response or the lowest confidence "old" or the lowest confidence "new" response for the ratings ROC and the probability of an "old remember" judgment or an "old know" judgment or a "new know" judgment on RK ROC. Because these are cumulative ROCs, the fourth point must correspond to an HR–FAR of 1.0, 1.0; therefore it is not plotted, as is the standard convention. The slope of the z-transformed REM ROC is .80. The fit is based on a Monte Carlo simulation of 10,000 subjects on 30-item lists. The parameters used for the REM fit are $u^* = .28$, $w = 20$, $t = 1$, $c = .70$, $g = .42$, and criteria = .30 (remember–know new, .65 (old–new), and 1.60 (remember–know old).

(REM) ratings (Shiffrin & Steyvers, 1997) and RK models (Malmberg, Zeelenberg, & Shiffrin, 2004).[1] The REM global-matching RK model assumes a yes–no criterion and a stricter RK criterion. To determine whether or not the stimulus was studied, its familiarity is compared to a yes–no criterion. If the initial response is "yes," and the familiarity also exceeds the RK criterion, the second response is "remember"; otherwise, the second response is "know." The ratings task is performed in exactly the same way, except that there are more criteria. Thus, the REM ratings and RK models predict the same ROC and the same zROC. Details of the fit of the single-process model to the data, which is quite accurate, are described in Figure 2. None of the estimated probabilities deviate by more than .015 from the data.

According to the single-process model, the mean slopes of ratings and RK zROCs should be similar and less than

1.0. These predictions were confirmed. Linear regression was performed on the z-transformed HR–FAR pairs for each subject. The mean slopes of the three-point ratings and RK zROCs were .80 and .72, respectively; both are reliably less than 1.0 [$t(37) = 4.49$, and $t(37) = 6.95$, both ps < .0005], but they are not reliably different from each other ($p > .26$). A maximum likelihood estimate of the slopes using the ROCKIT software program produced estimates of .74 and .75 for the RK and ratings zROCs, respectively. The slope of the mean REM zROC was .80.

Taken together, the HR–FAR pairs, the fit of the single-process model, the zROC analyses based on the individual subject data, and the maximum likelihood analysis all lead to the conclusion that the ratings and RK ROCs are indistinguishable. To determine whether or not the averaging method leads to a different conclusion, we performed an additional analysis. The slope of the ratings zROC generated from the mean response probabilities is .81, a value that is slightly greater than the .80 mean slope of the individual ratings zROCs. Thus, the averaging method had almost no effect on the slope of the ratings zROC.

In stark contrast, the slope of the averaged two-point RK zROC is .88, which is over 22% greater than the conventional estimate of .72. Thus, the averaging method substantially overestimated the mean slope of the single-subject RK zROCs. In fact, the group mean estimate reversed the numerical relationship between the slopes of the RK and ratings zROCs, producing a finding consistent with Rotello et al.'s (2004) finding. An overestimate of 22% produces slope estimates greater than 1.0 for zROCs with slopes greater than .82, a value that falls within the typical range.

The slope of the two-point ratings zROC based on averaged HRs and FARs also leads to an overestimate of the actual slope of the three-point zROC (.83 vs. .80), although the overestimate was not as great as for the RK zROC. The averaged three-point RK zROC produced a slope of .69, which is slightly less than the actual three-point slope of .72, but far less than the slope of the averaged two-point RK slope of .88. The averaged three-point ratings zROC produced a slope of .81, almost identical to the actual slope of .80. Thus, the nonlinearity that is introduced to the zROC by averaging tends to overestimate the two-point slopes (here by an average of ~13%), but has little or no effect on the three-point slope estimates (here by an average of ~3%). An overestimate of the slope of the RK zROC as a result of averaging provides a plausible explanation of Rotello et al.'s (2004) findings.

Here, we have focused on the relation between the ratings and two-point RK zROCs, but it is worth noting that Rotello et al.'s (2004) finding is not unique. For instance, Hintzman (2004) and Proctor (1977) reported that the slope of the ratings zROC is significantly less than the slope of the zROC generated from judgments of frequency (JOFs). It is beyond the scope of the present short article to delve into this issue in detail, but we simply note that like the two-point RK zROC, the JOF zROC is mostly confined to the strict portion of zROC space, because there are far more points associated with frequency judgments

>1. If for some reason the *z*ROC is concave-down, then one would expect the slope of the JOF *z*ROC to be greater than the slope of the ratings ROC. In the next section, we demonstrate how noisy decision processes can affect the form of the ROC.

Decision-Noise Models of Nonlinear *z*ROCs

Whereas averaging almost certainly contributed to Ro-tello et al.'s (2004) findings, it is also possible that some of the ratings *z*ROCs included in their meta-analysis might have actually been nonlinear prior to averaging. It is commonly stated that the recognition memory ratings *z*ROC is linear (see, e.g., Glanzer, Hilford, & Kim, 2004), but the evidence indicates that its form is highly variable. For instance, numerous ratings *z*ROCs have been reported that are inverted U-shaped. Heathcote (2003) reported 28 *z*ROCs, of which about 20 are slightly inverted U-shaped, and several nonlinear, single-subject ratings *z*ROCs were reported by Ratcliff, McKoon, and Tindall (1994). Non-linear *z*ROCs are sometimes attributed to "noisy" data. If this were the case, they should be randomly shaped. However, the forms of Heathcote's and Ratcliff et al.'s nonlin-ear *z*ROCs appear concave-up, concave-down, or nearly linear, and this suggests a systematic influence.

By definition, slopes calculated at different points on a nonlinear *z*ROC will be different. At the strict end of an inverted U-shaped *z*ROC, for instance, the slopes will tend to be greater than slopes obtained from the entire *z*ROC. However, inverted U-shaped *z*ROCs have been ob-served that could not have been distorted by averaging. Importantly, the averaging account cannot predict nonlin-ear single-subject *z*ROCs (see, e.g., Ratcliff et al., 1994) unless some form of within-subjects averaging was em-ployed, and hence the question remains: Why are so many ratings *z*ROCs inverted U-shaped?

Suboptimal decision processes can affect the form of ROCs (Broadbent, 1966; Erdfelder & Buchner, 1998; Krantz, 1969; Larkin, 1965; Malmberg, 2002; Nachmias & Steinman, 1963; Ratcliff et al., 1994; Wixted & Stretch, 2004), and we refer to these as *decision-noise models*. Malmberg (2002) showed how suboptimal decision strat-egies can affect the shape of the ratings ROC when they are generated from threshold models, and Ratcliff et al. (1994) demonstrated that adding a small amount of deci-sion noise to a continuous-state model produces an in-verted U-shaped *z*ROC.

Here, we focus on the influence that a noisy decision process has on the form of the continuous-state ratings *z*ROC. Figure 3 plots two REM ROCs and two REM *z*ROCs. An uncontaminated REM function (i.e., no deci-sion noise has been introduced) is represented with white circles. The function represented with black squares was generated using Ratcliff et al.'s (1994) method for intro-ducing decision noise. This is accomplished by first adding $n/100$ to each of the HRs and FARs on an ROC obtained from the uncontaminated REM model, and then normal-izing these probabilities by dividing each by $1.0 + n/100$ (Ratcliff et al., 1994). In this case, $n = 10$. The Ratcliff et al. decision-noise model produces a shifted ROC and an inverted U-shaped *z*ROC. If subjects sometimes adopt this decision strategy, this would account for at least some of the inverted U-shaped *z*ROCs that have been reported.

On the other hand, subjects are often instructed to use all available ratings or to evenly distribute responses to ratings in order to ensure that a full HR–FAR datum is obtained for each level of confidence. In contrast, the as-sumption underlying the Ratcliff et al. (1994) model is that noise is added only to the strictest rating. Thus, the Ratcliff et al. model might be used only infrequently. Nev-ertheless, the instructions to evenly distribute responses to
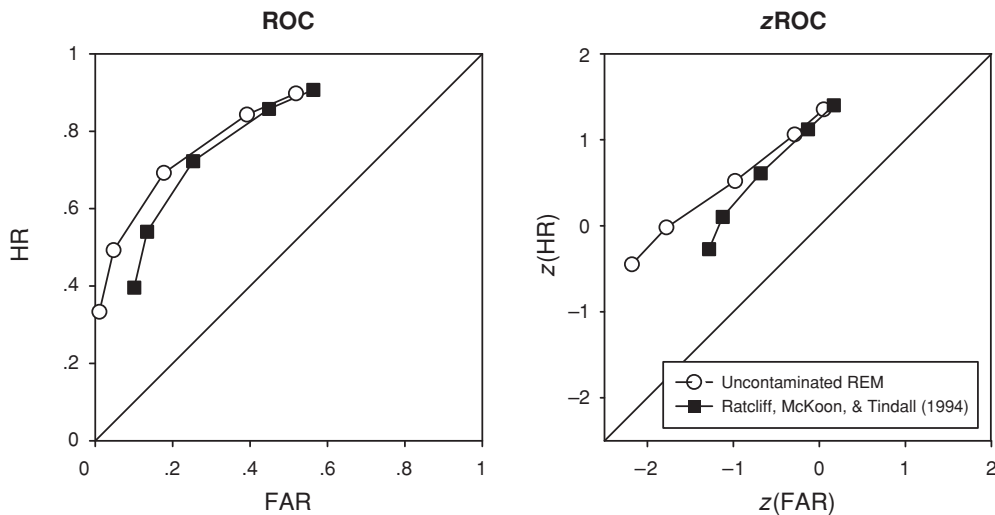

Figure 3. ROCs and *z*ROCs generated from the uncontaminated REM model and the Ratcliff, McKoon, and Tindall (1994) decision-noise model implemented in REM. The decision strategy described by Rat-cliff et al. produces a positive shift in the ROC and a concave-down *z*ROC. HR, hit rate; FAR, false alarm rate.

ratings might cause subjects to adopt a different, noisy decision strategy, one that assumes that noise is distributed across the ROC. Subjects make a valid decision with probability n; otherwise, they choose a rating randomly. We refer to this model as a *random decision-noise model*.

Figure 4 plots three ROCs and the corresponding three zROCs, each distorted by 50% random decision noise. The ROCs and zROCs assume the same set of criteria: 0.1, 0.37, 2.0, 3.0, and 12.0, but each ROC represents a different level of performance (accomplished by varying the REM u^* parameter at three levels: .04, .07, and .10). The shape of the random-noise zROC is influenced by performance level. At the lowest level of performance ($u^* = .04$), the zROC is inverted U-shaped. There are very few responses assigned to the extreme ratings; hence, this is where the impact is greatest from the random assignment of noise, driving the ends of the zROCs down toward chance. At the highest level of performance ($u^* = .10$), the zROC is U-shaped. Now the extreme ratings are receiving relatively more valid responses, and some of these are being introduced in the form of noise to the intermediate ratings, driving these points down toward chance levels. At the middle level of performance ($u^* = .07$), the zROC appears to be linear.

The random decision-noise model predicts a distinctive pattern in the form of the zROC as a function of overall performance. When a random decision-noise model is assumed, changes in sensitivity result in changes in the form of the zROC. When performance of the average subject is relatively high, there will be a tendency to observe U-shaped zROCs. When average performance is relatively low, there will be a tendency to observe inverted U-shaped zROCs. It is tempting to conclude that such a pattern of data reflects a heightened contribution of recollection or some threshold-like process to recognition when performance is relatively high (Yonelinas, 2002). However, it is also possible that subjects are using a noisy decision strategy.

REFERENCES

ATKINSON, R. C., & JUOLA, J. F. (1974). Search and decision processes in recognition memory. In D. H. Krantz, R. C. Atkinson, R. D. Luce, and P. Suppes (Eds.), *Contemporary developments in mathematical psychology: Vol. 1. Learning, memory, and thinking* (pp. 243-293). San Francisco: Freeman.

BROADBENT, D. E. (1966). Two-state threshold model and rating-scale experiments. *Journal of the Acoustical Society of America*, **40**, 244-245.

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.

EGAN, J. P. (1958). *Recognition memory and the operating characteristic*. Bloomington: Indiana University Hearing and Communication Laboratory (AFCRC-TN-58-51).

ERDFELDER, E., & BUCHNER, A. (1998). Process-dissociation measurement models: Threshold theory or detection theory? *Journal of Experimental Psychology: General*, **127**, 83-96.

ESTES, W. K. (1956). The problem of inference from curves based on group data. *Psychological Bulletin*, **53**, 134-140.

ESTES, W. K. (2002). Traps in the route to models of memory and decision. *Psychonomic Bulletin & Review*, **9**, 3-25.

FRANCIS, W. N., & KUČERA, H. (1982). *Frequency analysis of English usage: Lexicon and grammar*. Boston: Houghton Mifflin.

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

GLANZER, M., HILFORD, A., & KIM, K. (2004). Six regularities of source recognition. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **30**, 1176-1195.

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

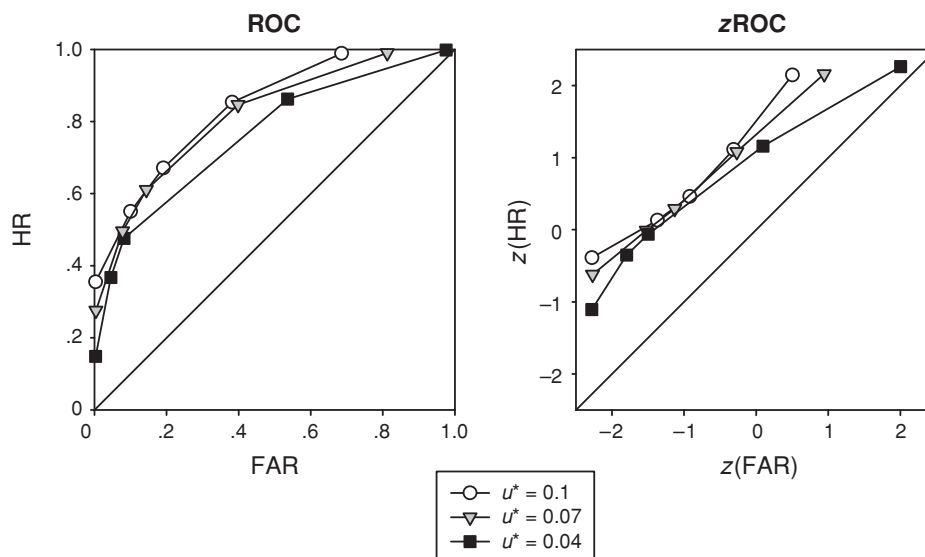HEATHCOTE, A. (2003). Item recognition memory and the receiver oper-

Figure 4. ROCs and zROCs generated by the random decision-noise model. The random decision-noise strategy affects the shape of the zROC. u^* is an REM-encoding parameter that determines how much information is stored on a study trial. As encoding improves, the form of the zROC transitions from concave-down to U-shaped. HR, hit rate; FAR, false alarm rate.

ating characteristic. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **29**, 1210-1230.

HINTZMAN, D. L. (2004). Judgment of frequency versus recognition confidence: Repetition and recursive reminding. *Memory & Cognition*, **32**, 336-350.

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

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

KRANTZ, D. H. (1969). Threshold theories of signal detection. *Psychological Review*, **76**, 308-324.

LARKIN, W. D. (1965). Rating scales in detection experiments. *Journal of the Acoustical Society of America*, **37**, 748-749.

MACMILLAN, N. A., & KAPLAN, H. L. (1985). Detection theory analysis of group data: Estimating sensitivity from average hit and false-alarm rates. *Psychological Bulletin*, **98**, 185-199.

MALMBERG, K. J. (2002). On the form of ROCs constructed from confidence ratings. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **28**, 380-387.

MALMBERG, K. J., HOLDEN, J. E., & SHIFFRIN, R. M. (2004). Modeling the effects of repetitions, similarity, and normative word frequency on old–new recognition and judgments of frequency. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **30**, 319-331.

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, & Cognition*, **30**, 540-549.

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

NACHMIAS, J., & STEINMAN, R. M. (1963). Study of absolute visual detection by the rating-scale method. *Journal of the Optical Society of America*, **53**, 1206-1213.

PROCTOR, R. W. (1977). The relationship of frequency judgments to recognition: Facilitation of recognition and comparison to recognition-confidence judgments. *Journal of Experimental Psychology: Human Learning & Memory*, **3**, 679-689.

RATCLIFF, R., MCKOON, G., & TINDALL, M. (1994). Empirical generality of data from recognition memory receiver-operating characteristic functions and implications for the global memory models. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **20**, 763-785.

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, & Cognition*, **26**, 294-320.

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.

SHIFFRIN, R. M., & STEYVERS, M. (1997). A model for recognition memory: REM—retrieving effectively from memory. *Psychonomic Bulletin & Review*, **4**, 145-166.

TULVING, E. (1983). *Elements of episodic memory*. New York: Oxford University Press.

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

YONELINAS, A. P. (2002). The nature of recollection and familiarity: A review of 30 years of research. *Journal of Memory & Language*, **46**, 441-517.

NOTE

1. Readers may refer to the primary sources for detailed descriptions of the single-process models (Malmberg, Zeelenberg, & Shiffrin, 2004; Shiffrin & Steyvers, 1997).