The systematic discrepancy between A' for overall recognition and remembering: A dual-process account

IAN G. DOBBINS

Harvard Medical School, Charlestown, Massachusetts

Signal detection accounts of recognition assume that all item endorsements arise from the assessment of a single continuous indication of memory strength, even when subjects claim to categorically separate items accompanied by contextual recollection from those that are not (viz., remembering vs. knowing). Dissociations of these response types are held to occur because the former require a higher response criterion for item strength than does the latter. Meta-analytic and individual subject data suggest that when the A' metric is used, accuracy for remembering can systematically deviate from that of overall responding for individual subjects. This occurs because, unlike the symmetric and rigid receiver-operating characteristic (ROC) implied under A' , empirical ROCs are asymmetric and plastic. A dual-process model predicted that the magnitude of the deviation would vary as a systematic function of the proportion of overall recognition accompanied by subjective remember reports for individual subjects. The predictions were confirmed using multiple regression on Monte Carlo and experimental data sets and were also shown to generalize to the double equal-threshold, single high-threshold [i.e., $H - FA$; $(H - FA)/(1 - FA)$], and the equal variance signal detection d' corrections. The unequal variance signal detection model was also shown to mirror the data, but only under the post hoc assumption that every subject adopts a very similar remember criterion placement rule. The results demonstrate that the systematic failure of tightly constrained models of recognition constitutes valuable regression data for more complex models and simultaneously highlights why single-point measures of accuracy are unsuitable as summaries across conditions or groups. Furthermore, the results show that remember rates carry unique information regarding the underlying processes governing individual subject performance that cannot be gleaned from the overall hit and false alarm rates in isolation.

William James suggested as early as 1890 that, on the basis of subjective experience, at least two separable processes seemed to underlie the experience of recognition. An oft cited anecdote suggesting two recognition processes was later provided by Mandler (1980), who described the "butcher in the bus" phenomenon, where one steps onto a bus and immediately recognizes a particular face as having recently been seen but is, at least temporarily, completely unable to describe the contextual details that would situate the feeling of familiarity. Adopting a distinction later formalized by Endel Tulving in 1985, this might be characterized as *knowing* that the person has been previously seen, without *remembering* elements sufficient to specify the source or occasion. Tulving (1985) argued that remembering and knowing represented different forms of conscious awareness that reliably map onto separate underlying memory systems and, therefore, should be relatively easy to experimentally dissociate. Considerable research has

supported this hypothesis, demonstrating dissociations of remembering and knowing with the use of encoding and retrieval manipulations (for a review, see Rajaram & Roediger, 1997), electrophysiological measurements (Duzel, Yonelinas, Mangun, Heinze, & Tulving 1997), and functional magnetic resonance imaging (Henson, Rugg, Shallice, Josephs, & Dolan, 1999). These dissociations have been taken by many as clear evidence for the presence of at least two separable mnemonic processes underlying explicit recognition.

Despite the introspective appeal and experimental evidence supporting dual-process accounts of recognition, several signal detection theorists have argued that remember and know responses may be entirely accommodated within the equal variance signal detection model by simply assuming that subjects adopt a secondary, more stringent response criterion for providing remember reports (two-criterion signal detection model; e.g., Donaldson, 1996; Hirshman & Henzler, 1998; Hirshman & Master, 1997; Inoue & Bellezza, 1998). From this perspective, "remember responses represent *nothing more* than conservative yes responses" (italics added; Donaldson, 1996, p. 524). On the basis of this characterization Donaldson suggested that criterion-free estimates of accuracy (e.g., A' or d') should be equivalent for overall

This research was supported in part by the National Institutes of Health. I thank Russell Poldrack, Anthony Wagner, and Andrew Yonelinas for their reviews of an earlier version of this manuscript. Correspondence concerning this article should be addressed to I. G. Dobbins, MGH-NMR Center, Bldg. 149, 13th St., Mail Code 149(2301), Charlestown, MA 02129-2060 (e-mail: ian@nmr.mgh.harvard.edu).

recognition and recognition selectively accompanied by remember reports. However, when the A' metric was used during meta-analysis, Donaldson failed to find equivalence for recognition and remember accuracy. Instead, A' for remember responses was consistently *smaller* than A' for overall responding (60 of 72 cases, $z = 5.54$). Although one might conclude that such an outcome was a statistical anomaly, Gardiner and Gregg (1997) have provided a brief review that suggests that the effect occurs quite often in individual-subject data, particularly under circumstances in which very little remembering is predicted to occur (e.g., very shallow or brief encoding). This pattern clearly violates the predictions of an equal variance signal detection account of remember/know performance. However, most memory researchers are aware that, unlike the symmetric receiver-operating characteristic (ROC) implied under the equal variance signal detection model, empirical ROCs are in fact often highly asymmetric, with the left-hand portion of the ROC elevated with respect to the right-hand portion. This asymmetry has typically been accommodated either by assuming that the old-item distribution is in fact more variable than the new (e.g., Ratcliff, Sheu, & Gronlund, 1992) or by assuming that performance is governed by more than one process (e.g., Yonelinas, 1994). Given that the two-criterion equal variance signal detection model advanced by Donaldson has been proffered as evidence against the notion that remember/know performance is governed by two processes, it is important to determine whether the tendency of A' estimates for remember performance to deviate from those for overall recognition can be explained by either the more complex unequal variance signal detection model or the competing dual-process signal detection model. In the sections that follow, it is demonstrated graphically that the dual-process model makes straightforward predictions regarding both the direction and the magnitude of the A' discrepancy (and for other point estimates of accuracy) on the basis of the idea that subjective remember reports carry critical information regarding the shape of the underlying ROC that characterizes recognition performance for each individual subject. Such a prediction is directly at odds with the notion that remember reports constitute nothing more than conservative *yes* responses.

In the present brief report, the relative operating characteristics (ROCs) implied by A' , the dual-process signal detection model (Yonelinas, 1994), and the unequal variance signal detection model are examined in order to generate predictions regarding both the direction and the magnitude of the discrepancy between A' for overall recognition and A' for remembering [A' (recognition) – A' (remember): the A' discrepancy] for individual subjects. These predictions are then tested using multiple regression on experimental and Monte-Carlo-generated data sets. The implications of the results for the signal detection and dual-process accounts of remember/know performance are discussed, and the results are shown to also easily generalize to the double equal-threshold, single high-threshold [i.e., hits – false alarms or (hits – false alarms) / (1 – false

alarms)], and equal variance d' accuracy measures, which like A' , also imply rigid ROC surfaces. The utility of these single-point accuracy measures with respect to model testing is shown to arise from the high degree of constraint or *rigidity* of their implied ROCs. Under these rigid models, if one knows a subject's overall level of accuracy, any subset of performance must fall to the left of this point along a known fixed path projected by the implied ROC. Any deviation from this path tells us something about the quantitative manner in which the shape of the subject's actual ROC differed from that predicted under the rigid model. The primary goal of the present report is to determine the degree to which the dual-process model, the unequal variance signal detection model, or both models provide a natural *a priori* prediction regarding the direction and size of this anticipated deviation as a function of each subject's overall performance and remember rate.

Graphical ROC Analysis and Predictions

One of the most important tools employed in the study of recognition has been the relative (or receiver) operating characteristic (ROC) (for a review, see Macmillan & Creelman, 1991; Swets, 1986). The implied ROC of a given model is the graph of the predicted cumulative relationship between the likelihood that an observer will correctly endorse an old item (termed a *hit*) and the likelihood that he or she will incorrectly endorse a new item (termed a *false alarm*) as a function of response confidence/criterion. The utility of the ROC lies in the fact that once a particular model of old/new discrimination is sufficiently specified, the possible shapes of the observable ROC functions that are consistent with the model are increasingly constrained; the repeated failure to find empirical data consistent with a model's predictions is grounds for modifying or abandoning the model. Figure 1A shows three separate implied ROCs, or isosensitivity curves, for the A' accuracy metric as performance improves from an A' of .75, to an A' of .95.¹ The shape of the three curves remains constant even as overall performance improves.² That is, the ROC for A' can be imagined as a rigid surface that is slid upward along the minor diagonal as performance improves. Subjects with a given accuracy are free to move along this surface as a function of their adopted response criterion, but their performance should not reliably deviate above or below the surface, regardless of the nature of their subjective report. This fact, combined with the signal detection assumption that remember responses merely reflect very conservative *yes* responses, should enable one to precisely predict a given subject's correct remember rate as a function of his or her overall A' . For example, given a subject with an A' of .85 for overall recognition (middle ROC, Figure 1A), one would predict that his or her observed remember rate should be about .40 if the criterion is set so high as to preclude incorrect remember responses. In short, the A' model perfectly constrains the possible correct remember rates as a function of overall performance and the supposed remember re-

sponse criterion because the surface of the ROC is rigid. Violations of this prediction tell us something about the manner in which the empirical ROC deviates from the well-characterized ROC of A' . In this manner, A' can serve as a known benchmark or point of departure enabling analysis at the individual-subjects level.

Dual-process model predictions. The high degree of shape invariance of the A' ROC stands in marked contrast to the ROCs implied by the dual-process signal detection model (Yonelinas, 1994), which are illustrated in Figures 1B and 1C. The dual-process signal detection model assumes that successful performance is governed by two underlying recognition processes: recollection of contextual information, which is modeled as a threshold occurrence, and a familiarity discrimination process consistent with the equal variance signal detection model underlying the d' accuracy index. Both sources of information independently combine to influence the observed hit rate and this is represented by the equation

$$P(\text{hit}) = R_o + (1 - R_o) * \phi(d'/2 - c),$$

where R_o is the probability that an old item evokes recollection and is assumed to reliably covary with each subject's overt remember rate (see Yonelinas, Dobbins, Szymanski, Dhaliwal, & King, 1996, for demonstrations), ϕ is the cumulative normal distribution function, d' is the distance separating the unit normal old- and new-item familiarity distributions assumed under equal variance signal detection, and c is the familiarity criterion for responding *old*. Thus, an observer is expected to succeed if he or she experiences recollection (R_o) or, in the absence of recollection ($1 - R_o$), the test item is sufficiently familiar to ex-

ceed the common familiarity response criterion at a given level of confidence (see Yonelinas et al., 1996, for more detail).

Figure 1B shows the behavior of the dual-process ROCs as recollection increases while the overall recognition performance is held at a constant ($A' = .775$). Unlike the rigidity of the A' ROCs, the dual-process ROCs noticeably flatten, and the intercept elevates, as recollection increases. Under the assumptions of the dual-process model, the points at which the ROCs intercept the y-axis are those that should closely predict the overt correct remember rates of each subject. If one were to use the A' accuracy measure to compare overall recognition and remembering for each ROC and the dual-process model was in fact correct, not only should there be discrepancies in the values (because the flexible shapes of the dual-process ROCs do not match the shape implied by A'), but the magnitude of this difference should systematically vary as a function of the proportion of the subject's total recognition determined by remembering. That is, if one were to imagine grabbing and holding the ROC fixed at the midpoint corresponding to overall performance, the left-hand portion is still free to slide up or down, depending on the subject's remember rate (i.e., the ROC is *plastic*, as opposed to *rigid*). For example, in Figure 1B, the horizontal arrow marks the fixed y-intercept of .10 assumed under A' when overall recognition corresponds to a hit rate of .69 and a false alarm rate of .31 ($A' = .775$). In contrast, the dual-process y-intercepts are free to vary around this location, even if the overall recognition rate remains the same for two different observers. To illustrate this, I have fixed the overall recognition rate to the same point on the

Figure 1A. The rigid surface of the A' receiver-operating characteristic.

Figure 1B. The predicted dual-process effects of increasing the recollection process from zero to .55 while holding overall performance constant. The overall A' is expected to remain constant, whereas the A' for remember responses is predicted to systematically increase with recollection.

minor diagonal in Figure 1B ($H = .69$, $FA = .31$) for four dual-process ROCs in order to demonstrate their behavior as recollection increases while overall recognition remains at a constant location on the probability plot. The dotted line in Figure 1B represents the projected surface of the A' ROC. There are two important points to note about the generation of the dual-process ROCs under the constraint that overall recognition is held fixed to a constant location. First, all four ROCs share the same overall signal detection criterion for responding *old*. This is true because all four ROCs share the same overall false alarm rate, which for simplicity will initially be assumed to be driven entirely by item familiarity; this in turn means that the familiarity criterion is equivalent.³ Second, because the overall hit rate is instead determined by two processes, increases in recollection must be accompanied by decreases in familiarity (i.e., d') in order to maintain the same overall performance hit rate across the ROCs. The lowest curve demonstrates the predicted performance when there is very little recollection ($R_o = .01$). The d' required to maintain an overall recognition rate of (.69, .31) is approximately .98 (see Equation 1). Importantly, this ROC intercepts the y-axis below the intercept predicted by A' . Thus, the dual-process model predicts that given this particular level of overall performance, subjects with very low remember rates will have A' scores for remember responses that are lower than what would be expected on the basis of A' for overall recognition. If, however, we suppose that recollection occurs on 20% of the trials and overall recognition remains constrained to the same values, then by Equation 1, the y-intercept is .20,

and d' must be reduced to .79 to maintain the same overall recognition rate. Importantly, the y-intercept and the left-hand portion of the ROC now lie above that predicted by A' . This pattern continues until a recollection rate of approximately .55, after which it becomes impossible to lower d' any further in order to maintain same overall recognition rate (i.e., d' would have to be less than 0).

In summary, the dual-process model predicts that within a group of observers with similar overall recognition scores, those who infrequently report remembering will tend to have A' remember values smaller than what would be suggested by their overall recognition A' (i.e., A' discrepancy is positive). This is consistent with previous findings that such discrepancies tend to occur in studies using very minimal encoding conditions. The new proposal advanced here is that this difference will *systematically* decrease as a function of remembering until, at some point, the remember rate is high enough to elevate the intercept of the ROC above that implied by A' [i.e., A' (overall) – A' (remember) will be negative]. This can be briefly summarized by stating that the relationship between remembering and the A' discrepancy is negative, given that overall recognition performance is held constant.

In contrast to the predicted effects of the remember rate on the difference in A' values, Figure 1C shows the ROCs implied by the dual-process model when, instead, recollection is held fixed and overall performance systematically increases across subjects. Because the overall recognition hit rate is codetermined by both recollection and familiarity, increases in overall performance while remembering is held constant require increases in the familiarity

discrimination process across subjects (i.e., increases in d'). Under these conditions, the overall performance level will increase, but the point at which the ROCs intersect the y-axis will remain fixed near the constant remember rate. Here, the relationship between overall performance and the A' discrepancy will be in the opposite direction to that predicted for the relation between remembering and the A' discrepancy in Figure 2. That is, as overall performance becomes proportionally more determined by increased familiarity discrimination, the ROCs will become more bowed than is predicted under the A' model, because the curvature of the ROC will increase with increases in d' . Given this, the difference between A' for overall recognition minus that for A' remembering will also systematically increase. This can be summarized by stating that the relationship between overall recognition and the A' discrepancy is positive, given that the level of remembering is held constant across subjects.

In summary, the dual-process model predicts opposite effects of remember rates and overall performance rates on the observed differences between A' for overall recognition and A' for remembering. With all other things being equal (viz., familiarity and overall performance level), increases in recollection will result in an elevation of the left-hand portion of the ROC and a flattening of its shape. In contrast, with all other things being equal (viz., recollection), increases in familiarity will result in an elevation near the center of the ROC and increased bowing of its shape. Of course, experimentally it would be impos-

sible to actually hold either overall recognition or remember rates constant across different subjects while manipulating the other, since encoding and retrieval manipulations often typically affect both. However, using multiple regression analysis, one can assess whether the overt correct remember rate and the overall recognition rate have statistically *unique and opposite* effects on the measured discrepancy of A' overall and A' for remembering for individual subjects (Pedhazur, 1997). That is, multiple regression analysis allows one to examine the expected change in a dependent variable associated with a unit change in an independent variable while holding constant the effects of the other independent variables in the model. In the present context, one can examine the unique contributions of the remember rate and overall performance rate in the prediction of the A' discrepancy. The dual-process prediction is that the regression coefficient for the remembering will contribute significantly and negatively to the A' discrepancy, whereas that of the overall performance variable will contribute significantly and positively to this difference. This prediction is first tested on experimental data and then further validated by using Monte-Carlo-generated data that assume a dual-process model.

Unequal variance signal detection model. As with the dual-process model, the unequal variance signal detection model also produces asymmetric and plastic ROCs. The asymmetry arises because, under the model, it is assumed that studying a random selection of items systematically increases their variability in strength values with

Figure 1C. The predicted dual-process effects of increasing the familiarity process while holding the recollection rate constant. The remember A' is expected to remain constant, whereas the A' for overall responding is predicted to systematically increase.

respect to a random sampling of novel items. This is illustrated in Figure 1D. As one “flattens” the old-item familiarity distribution with respect to the new, this elevates the left-hand portion of the ROC and imparts a slight “S” shape to the curve. The reason for this becomes apparent when one looks at the distributions in the lower corner of Figure 1D. Starting at the right-hand side of the distributions, each unit moved leftward on the scale initially results in a proportionally much greater increase in the hit rate than in the false alarm rate as the old-item variance is increased; thus, the rise in cumulative hit rates on the ROC is relatively sudden, in comparison with the rise in false alarm rates. However, as one begins to encroach onto the new-item distribution, this pattern reverses, and the rise in cumulative hit rates now begins to slow with respect to that of the false alarm rate. Thus, the left-hand portion of the ROC elevates as the variance of the old-item distribution increases, and the right-hand portion depresses. The key thing to note at this point, however, is that there is no *a priori* reason to suspect that increasing the variance of the old-item distribution should affect not only the shape of the ROC, but also the predicted remember rates of individual subjects. This is because the placement of criteria within signal detection is typically assumed to be largely idiosyncratic across individuals. The only additional constraint added to the two-criterion signal detection model proposed by Donaldson (1996) was that the remember criterion should fall to the right of the old/new location. If its placement is also idiosyncratic and highly variable across individuals, the observed remember rate of each individual, like the observed hit rate, will be largely unrelated to the old-item variance and, hence, the shape of the underlying ROC. This means that the actual observed remember rate should not predict the discrepancy in A' values across overall and remember performance. In contrast, if each subject adopts a very similar rule for the placement of the remember criterion, a relation between remember rates and old-item variance might emerge. For example, if all the subjects adopted the rule to place their remember criterion 1.5 new-item standard deviation units to the right of their old/new criterion, then all other things being equal, a subject with a higher old-item variance would tend to have more of the old-item distribution fall to the right of his or her remember criterion. Thus, the remember rate would now be indicative of the old-item variance, which in turn specifies the elevation/asymmetry of the ROC and, therefore, predicts the degree of discrepancy across A' for overall performance and remembering. Returning to Figure 1D, as the old-item variance increases, the ROC will climb above that predicted by A' . At moderate old-item distribution variances, one should expect that the A' for remembering will be higher than that for overall performance. However, this prediction is not absolute, because unlike the dual-process and A' ROCs, the unequal variance signal detection ROC must intercept the origin regardless of the old-item variance and d' distance. Given

this, there will always be some point at which every unequal variance ROC crosses the A' ROC on its way toward the origin. In practice, however, this should rarely, if ever, occur for moderate old-item variances, because it requires an exceedingly conservative criterion for remember report in order to find locations on the unequal variance ROC below those implied by A' . For example, with an A' of .775, the A' ROC intercepts the y-axis at a hit rate of 10%. The unequal variance model, in which the old-item standard deviation is twice that of the new (top curve, Figure 1D), will not fall below this ROC until the false alarm rate is well below a hundredth of a percent. Given the typical numbers of trials used in recognition experiments and the high standard error of z scores at these extremes, this in all practicality constitutes a false alarm rate of zero for which d' is undefined.

The predictions above relating old-item variance and the A' discrepancy will be more systematically examined in the Monte Carlo section that follows. However, two things are worth noting about assuming a stereotypic remember criterion placement rule across subjects. First, it is perhaps questionable to assume that individuals with different memory capacities and subjective experiences (owing to both the distance and the variance of the old-item distribution) would uniformly adopt a roughly identical rule for placing their remember criterion, especially since they do not appear to do so for their initial old/new criterion placement. Second, if one does assume a very similar rule across subjects, the remember rate cannot now be viewed as being nothing more than confident yes responses, because it will now carry important information regarding the relative variance of the old-item distribution with respect to the new for each subject.

Multiple Regression on Experimental Data

The experimental data analyzed in this section were drawn from three unrelated studies of recognition. They are similar in that each used a relatively shallow encoding task and required subjects to make remember/know discriminations during single-item recognition testing. The methods unique to each are summarized in the Appendix and in the original research reports. This data set has also been used in an unrelated investigation of the relationship between remembering and committing false alarms (Dobbins, Khoe, Yonelinas, & Kroll, 2000).

Multiple regression was performed on the aggregate data and each data set, using the correct remember rate and the A' for overall responding as predictors of the difference between the overall A' and the A' for remembering across subjects (the A' discrepancy).⁴ The A' for overall recognition was calculated by using each subject's overall hit and false alarm rates, and the A' for remembering was calculated by using the associated hit and false alarm rates for the remember responses of each subject. For each experiment, the data were screened for outliers, using Cook's distance (Pedhazur, 1997). This resulted in the removal of 1 subject from Experiment 1 whose Cook's D

Figure 1D. The behavior of the unequal variance signal detection model. The value of A' remember will systematically increase with respect to A' for overall performance as the old-item variance increases.

was .855. The descriptive statistics and regression results are presented in Tables 1 and 2. Overall, the dual-process predictions were confirmed. In the aggregate data and two of the individual data sets, the subjects' remember rates were significantly negatively associated with the A' discrepancy, whereas in contrast, the overall recognition A' (that is, overall performance with the effects of remembering statistically controlled) was positively associated with this difference. It is perhaps worth emphasizing that without the tool of multiple regression, it would have been difficult to determine the unique contribution of remembering to the A' discrepancy, because individual differences in remembering are conflated with differences in overall performance. Multiple regression enables one to parse this relation and test the graphical predictions that the two should have opposite effects on this discrepancy. Of course, if it were the case that both overall performance and remembering were driven by the same underlying single factor, considering the role of remembering apart from overall performance would have yielded little if any gain in predicting the discrepancy. That is, the predicted regression weight for remembering would have been zero, and there would be no increase in the variance accounted for (R^2) as a result of its inclusion (i.e., redundancy). This stands in contrast to the actual aggregate results, where the regression weight was significantly negative and the inclusion of remembering as a variable apart from overall performance resulted in an 11% in-

crease in the variance explained ($p < .001$; for a related discussion of redundancy and suppression in individual recognition data, see Dobbins et al., 2000).

However, in Experiment 2, the remember rate did not significantly predict the A' discrepancy. The methods for this experiment and the descriptive data on the remember rates suggest that this null result occurred because of a lack of observed remember reports. The prediction regarding the negative relationship between remember reports and the A' discrepancy is necessarily conditional upon having a sufficient proportion and range of remembering in the experiment, and in Experiment 2, both the range and the standard deviation of remember reports were almost halved, in comparison with Experiments 1 and 3. This is not surprising, given that the experiment employed the most "shallow" encoding of the three (see the Appendix). Thus, it appears that there was an insufficient amount of remembering present in Experiment 2 to form a reliable relation, and therefore, none would be expected under the dual-process model.

In the next section, I provide further evidence for the predictions of the dual-process model, using Monte Carlo simulations that verify that dual-process assumptions/mechanisms are capable of producing the pattern of regression effects noted above. The unequal variance signal detection model is also simulated, in an attempt to determine whether there is a method of remember criterion placement that mimics the regression results. Overall,

Table 1
Means and Standard Deviations for Experiments and Simulations

	Hits		False Alarms		A' Total		Correct Remembers		Incorrect Remembers	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
Experiment 1	.69	.12	.33	.18	.76	.08	.39	.13	.05	.06
Experiment 2	.51	.19	.30	.17	.69	.07	.18	.11	.05	.05
Experiment 3	.70	.13	.21	.11	.83	.05	.33	.17	.03	.03
Total experimental data	.63	.17	.28	.16	.76	.09	.30	.17	.04	.05
Dual-Process Simulation 1	.65	.15	.30	.19	.76	.08	.32	.16	.06	.06
Dual-Process Simulation 2	.56	.16	.27	.18	.74	.08	.16	.08	.03	.03
Unequal Variance Simulation 1	.62	.16	.27	.19	.77	.08	.34	.18	.07	.12
Unequal Variance Simulation 2	.62	.16	.28	.19	.77	.08	.32	.14	.03	.06

Note—Dual-Process Simulation 1 modeled the correct remember rate and standard deviation of the aggregate data, whereas Dual-Process Simulation 2 halved these values. Unequal Variance Simulation 1 placed the remember criterion randomly, whereas Unequal Variance Simulation 2 elevated the remember criterion a fixed amount with respect to the old/new criterion location. See the text for further details. Sample size for each experiment was 24, and for each simulation it was 7,200.

the above regression results clearly demonstrate that the remember rate and the overall performance level have unique and opposite effects on the predicted A' discrepancy and, therefore, each contains unique information with respect to the A' discrepancy.

Simulating the A' Discrepancy Effect With the Dual-Process Model

The purpose of this section is twofold. The first is to see whether the pattern in the aggregate data could be specifically modeled by using dual-process parameters sampled in a manner consistent with the descriptive statistics in Table 1. The second goal was to examine whether the failure of the correct remember rate to significantly predict the A' discrepancy in Experiment 2 resulted because insufficient remembering occurred during the experiment.

The first simulation was conducted by using values similar to the aggregate data to center the sampling processes. First, each of 72 artificial subjects was given a familiarity d' randomly sampled from a normal distribution, with a mean of .729 and a standard deviation of .48. These values were arrived at by solving the dual-process equation for each subject [$P(\text{hit}) = R_o + (1 - R_o) * F$] in order to find the value for F , the hit rate associated with the familiarity process. A similar equation was solved in order to find the false alarm rate associated with the familiarity process [$P(\text{fa}) = R_n + (1 - R_n) * F$], where R_n is the incorrect remember rate for each subject. The average of the familiarity d' when this method was used was found to be .729. After a d' value was sampled, a familiarity response criterion was randomly sampled for each subject by sampling familiarity hit rates from a normal distribution centered on .48, with a standard deviation of .19 (the mean and standard deviation of the derived hit rates above). With the use of this proportion to fix the old/new response criterion, the associated familiarity-driven false alarm rate was then calculated for each artificial subject. Next, a threshold correct remember rate was randomly sampled from a normal distribution having the same mean and standard deviation as the aggregate data correct remember rates. The incorrect remember rate was assumed to be a function of two factors—the ability of the subject

to retrieve threshold contextual information, in which case it should be directly related to the correct remember rate, and the degree of inadvertent semantic overlap between some proportion of the target and distractor test list items, sufficient to trigger contextual recollection. The latter is assumed to be purposefully low in most recognition studies, but some degree of random and accidental overlap is probably unavoidable, given the large study/test lists used in detection experiments (e.g., cup vs. mug). For the present simulations, I assumed that some random proportion centered around 15% of the foil items may be sufficiently semantically related to trigger false remembering (e.g., accent vs. dialect), depending on the list randomization/counterbalancing for each subject, and this resulted in a false remember rate similar to that in the actual experimental data. The observed false remember rate was then calculated from the product of this overlap proportion and the artificial subject's correct remember rate. Following this, the hit and false alarm rates for the remember and familiarity process were independently combined to obtain the overall performance hit and false alarm data. Row 5 of Table 2 shows representative regression beta weights and t values obtained from a sample of 72 artificial subjects constructed in such a manner. The results replicated the pattern observed in the experimental data; the remember rate is significantly negatively related to the A' discrepancy, and the overall recognition rate is positively related to this discrepancy. When 100 samples of 72 subjects was drawn without replacement from a population of 7,200 constructed in such a manner, the effect replicated at the .001 alpha level for 97 of the samples and at the .01 alpha level for all of the samples.

Row 6 of Table 2 demonstrates the effect of restricting the range and level of remembering. The simulation was conducted identically to the one above, except that the correct remember rate and deviation were halved. Under these circumstances, which mirror the reductions observed in Experiment 2, there is a sizable reduction in the negative relationship observed between remember rates and the A' discrepancy. Random sampling of groups of 24 from this data set typically revealed no evidence of a relationship, and on the basis of this observation, one would not have

Table 2
Regression Results for Experiment and Simulation Data Where the Correct Remember
and Overall Performance Rates Were Used to Predict the Discrepancy in A'

	Beta Weights		<i>t</i> Values		<i>R</i> ²
	Remembers	<i>A'</i> Overall	Remembers	<i>A'</i> Overall	
<i>A'</i> Discrepancy, Experiment Data					
Experiment 1, <i>n</i> = 23	−.26	1.06	−3.40*	14.00†	.91
Experiment 2, <i>n</i> = 24	.10	.62	.51	3.36*	.45
Experiment 3, <i>n</i> = 24	−.86	1.10	−6.27†	7.96†	.76
Aggregate, <i>n</i> = 72	−.40	.96	−4.61†	10.98†	.64
<i>A'</i> Discrepancy, Dual-Process Simulation, <i>n</i> = 72					
Simulation 1					
Full remember rate	−.34	.92	−4.78†	12.84†	.71
Simulation 2					
Remember rate halved	−.02	.87	−.39	13.78†	.75
<i>A'</i> Discrepancy, Unequal Variance Signal Detection Simulation, <i>n</i> = 72					
Simulation 1					
Remember criterion free	−.07	.84	−1.01	12.11†	.68
Simulation 2					
Remember criterion fixed	−.30	.86	−6.34†	18.48†	.85
Other Single-Point Estimates, Experiment Data					
	<u>Remembers</u>	<u>Overall</u>	<u>Remembers</u>	<u>Overall</u>	
DET (H−FA)					
Aggregate, <i>n</i> = 72	−1.10	1.11	−25.21†	27.65†	.93
SHT(H − FA)/(1 − FA)					
Aggregate, <i>n</i> = 72	−1.02	1.25	−32.56†	39.92†	.96
Equal Variance <i>d'</i>					
Aggregate, <i>n</i> = 57	−.52	.71	−3.52†	4.79†	.30

Note—For rows 1–8, the dependent variable is the numeric difference between A' for overall and remember responses. For rows 9 and 10, the dependent variable is the difference in threshold corrected accuracy for overall and remember responses. For row 11, the dependent variable is the difference in equal variance model *d'* for overall and remember responses. The smaller *n* occurs because subjects that failed to commit false alarms during remember report were discarded (i.e. *d'* is undefined). DET, double equal threshold; SHT, single high threshold. **p* < .01. †*p* < .001.

expected to detect such a relationship in the actual data of Experiment 2. Overall, the simulation data bolster the dual-process interpretation of the empirical data by demonstrating that the contribution of two independent sources of information, with characteristics mirroring the actual remember and overall data, easily reproduces the pattern in the actual regression results.

Simulating the A' Discrepancy Effect With the Unequal Variance Signal Detection Model

As was discussed earlier, in order for the signal detection model to accommodate the data, each subject's remember rate must carry information with respect to the size of his or her old-item variance. It was presumed that this would not occur if the remember criterion placement was idiosyncratic across each subject. The first simulation was conducted to confirm this. A second simulation was conducted to demonstrate that the effect could be mirrored by the unequal variance signal detection model, provided each subject employs a very similar rule and distance for the placement of the remember criterion. As with the dual-process simulations, values similar to the aggregate experimental data were used to center the sampling processes. Each artificial subject was given a *d'* drawn from a normal distribution with a mean of 1.05 and standard deviation of .48. Following this, they were given an old-item standard deviation drawn from a uniform distribution between 1 and 2. An overall hit rate, sampled in a manner matching

the actual distribution of hit rates, was used to fix the old/new criterion location and determine the overall false alarm rate. The remember criterion was then determined by assuming that the subjects would place the criterion to the right of the old/new. The distance of this remember criterion was determined by sampling from a uniform distribution between 1 and 3 standard deviations greater than the old/new criterion position. By using this criterion, the correct and incorrect remember rate were then calculated. If these values fell outside of the range of the actual data, the remember criterion was resampled. As was expected, this approach typically did not yield data that matched the actual experimental regression results. The relevant means and regression results for 7,200 subjects are shown in Tables 1 and 2. When 100 samples of size 72 were drawn from this population without replacement and entered into regression, 93% failed to demonstrate the relation between remember rates and the A' discrepancy at the .001 significance level. The failure to capture a relationship clearly resulted from the largely random placement of the remember criterion with respect to the location of the old/new criterion. Next, the unequal variance model with an identical placement rule across subjects was examined. This simulation was identical to that above, except that for every subject, the remember criterion was placed 1.5 standard deviations above the old/new location. In contrast to the preceding model, this model successfully mirrored the regression pattern shown in the experimental

data. Under these conditions, the remember rate was negatively associated with the A' discrepancy, and the values of samples of subjects drawn from a population of 7,200 were quite similar to those provided by the dual-process model (see Tables 1 and 2). When the overall level of accuracy is statistically controlled and subjects employ an identical (or highly similar) rule for placing the remember criterion with respect to the old/new location, those subjects with larger old-item variances will have greater remember rates.

DISCUSSION

The present study demonstrates that the previously observed discrepancies between A' for overall and remember responding are not anomalous but are, in fact, quite systematically related to the level of observed remembering and are predicted to occur under the dual-process signal detection model of Yonelinas (1994) in a straightforward manner. These A' discrepancies arise because the rigid shape of the ROC implied for A' is inconsistent with empirical data. Visual inspection of theoretical dual-process ROCs suggested that the discrepancy should be systematic and opposite for overall recognition performance and observed correct remember rates, and this was the case in both empirical and Monte Carlo data. In the case of remember rates, the regression results demonstrate that, with the effect of overall performance held constant, the level of reported remembering shares a systematic relation with deviations in the shape of the ROC of individual subjects; as the remember rate increases, the difference between A' for overall and remember performance moves in the negative direction of the number line. Because the surface implied by A' is rigid, this means that the left-hand portion of the empirical ROC must be elevating in concert with the remember rate when the effects of overall performance level are controlled for. This is clearly inconsistent with signal detection accounts of remember/know performance, in which the remember rate is solely the result of a conservative criterion placement within an equal variance signal detection model (e.g., Donaldson, 1996; Hirshman & Henzler, 1998; Hirshman & Master, 1997; Inoue & Bellezza, 1998).

Under the unequal variance signal detection model, increasing the old-item strength variance with respect to the new has the effect of imparting an S-shape to the ROC and, more important, elevating the left-hand portion of the ROC, the latter effect being similar to increasing recollection under the dual-process model. This, in turn, will affect the A' estimates of overall and very conservative responding (assumed to reflect remember responses under signal detection) in a fashion similar to that predicted by the dual-process model. However, in order for the unequal variance signal detection model to accommodate the present findings, it is necessary to link the old-item variance and the correct remember rates of individuals. It should be noted that there is a critical difference in assumptions between the present results and previous signal detection proposals regarding criterion placement

(e.g., Stretch & Wixted, 1998). Typically, these theories revolve around the reasonable assumption that a given subject will try to maintain a fixed criterion rule or location across changing experimental conditions or manipulations—for example, when switching between test lists in which the studied items were encoded in a relatively “strong” or “weak” manner. That is, increases or decreases in the proportion of responses at a given confidence level are held to occur because the subject is maintaining a criterion placement rule or location across changing experimental conditions. In contrast, the present data suggest that, in order to mimic the results, the unequal variance signal detection model requires a criterion placement explanation that operates almost identically *across different subjects* despite the fact that the phenomenal experience of each greatly differs (i.e., each is encountering a different distribution of perceived item strengths). The present results suggest that if one is willing to assume that studying a list of items increases their variance with respect to a random sample of new items and that subjects will adopt a very similar rule for the positioning of their remember response criterion, a systematic relation between the observed remember rate and the shape of the ROC can be expected. From a signal detection standpoint, this finding would be important, since it means that the remember rate actually provides critical information with respect to the ratio of new- and old-item variances that is not available from the overall hit and false alarm rates alone.

In contrast to use of simple item strength as a basis for response, other, more complex potential methods for setting criteria have been proposed. For example, one of the criterion placement models evaluated in Stretch and Wixted (1998) is based on the idea that subjects use a likelihood ratio when placing response criteria. Applied to the present data, one could assert that although subjects have different old-item variances, each establishes a similar criterion based on the likelihood ratio of successful *remember* responding. Stretch and Wixted noted two potential objections to the use of such a rule for criterion placement. The first is that the derivation of a likelihood ratio requires that the subjects know the locations and mathematical forms that describe the shape of the distributions in question (see Hintzman, 1994, for a similar objection). In fact, the subjects would have to know this information before the start of the test in order to apply a likelihood criterion rule reliably. The second problem arises because, in the present data, we must assume unequal variances between old- and new-item distributions in order to account for the changes in shape of the ROC. Stretch and Wixted pointed out that under the unequal variance model, there are often two solutions along the strength dimension that result in the same likelihood value, resulting in some ambiguity as to which criterion should be selected to attain a particular ratio. These major concerns aside, Table 3 provides some illustrative values for remember rates and criterion locations under the assumption that different observers will adopt a similar likelihood of success when providing remember reports, despite hav-

ing differing old-item variances.⁵ These values were determined under the assumption that subjects will select the most conservative criterion solution to the likelihood evaluation whenever two solutions are possible. The table is not meant to be exhaustive, merely to illustrate that there is a very complex and often fine-grained relation between the criterion location, the remember rate, and the old-item variance under the particular likelihood ratios examined. The first column of the table shows the desired likelihood of success for a remember response, the second shows three potential standard deviations of the old-item distribution, the third is the location on the strength axis necessary to maintain that particular likelihood of successful responding (scaled in units of the new-item distribution), and the fourth and fifth columns show the predicted effect on the observed remember rates. It was assumed that all the subjects had the same d' value of 1. Looking at the first three rows of Table 3, it can be seen that in order to maintain a likelihood of success of 3 to 1, the remember criteria have to be increasingly positioned to the right as the old-item variance increases. As a result, adopting this particular likelihood results in very little change in the predicted remember rates across the different variances, although the overall level of correct remembering is actually similar to the experimental data. This restricted range of predicted remembering cannot accommodate the relations in the present findings and is noteworthy, given the considerable span of old-item variance examined (1 to 2.25 times larger than that of the new-item distribution). In the second block of the table, the subjects attempt to maintain a more conservative likelihood of success of 6 to 1. Although the overall level of correct remembering is now lower than that in the actual experimental data, there is now a noticeable—albeit slight, relative to the experiment data—increase in correct remembering as the old-item variance increases. However, when examining the criterion locations, it can be seen that their positions differ only in the second decimal place. This would seem to suggest an undue amount of precision on the part of the observer in either estimating item strengths per se or likelihood ratios based on item strengths. Finally, the third block of Table 3 shows the predictions under an even more conservative likelihood of 9 to 1. The most prominent feature of this block is that now the criteria must track leftward in order to maintain the same likelihood across different old-item variances.

Overall, Table 3 does not suggest a clear relation between likelihood values, old-item variances, and remember rates. At the lower likelihood ratio, there does not appear to be enough change in the predicted remember rates, in comparison with the old-item variance, to mirror the present experimental data. At higher likelihoods, there does appear to be a positive relation between old-item variance and remember rates, but only across a small range of remember rates ($\sim 10\%$), and the locations of the criteria are extremely close on the strength axis, so that subjects have to be able to parse the strength axis to the second decimal place in order to successfully place their criterion. These potential difficulties are compounded when one considers that we are, in theory, talking about

Table 3
Remember Rate as a Function of Likelihood Ratio and Old-Item Variance

Likelihood Ratio	σ_{old}	Strength Criterion	Correct Remember Rate	Incorrect Remember Rate
3:1	1.00	1.599	.275	.055
	1.25	1.727	.280	.042
	1.50	1.818	.293	.035
6:1	1.00	2.292	.098	.011
	1.25	2.239	.161	.013
	1.50	2.258	.201	.012
9:1	1.00	2.697	.045	.003
	1.25	2.510	.114	.006
	1.50	2.488	.160	.006

Note—Values based on a d' of 1.

completely different subjects; any moderate degree of variability in the level of skill of criterion placement/estimation across subjects or inconsistency in the adopted common likelihood value would preclude the emergence of a relation between old-item variance and remember rate across the range considered in Table 3. Of course, it may be the case that a different combination of d' , old-item variance, and adopted likelihood criterion can be found that mirrors the present experimental data. However, the explanatory value of such a model would have to be weighed against the assumption that subjects calculate likelihood ratios prior to responding and the need to arbitrarily assume that different subjects will adopt a similar remember likelihood ratio criterion. Such a model is clearly much more complex than the unequal variance strength model examined earlier and fails to offer any benefit as a result of this increased complexity.

A more pragmatic implication of the present study relates to the frequent use of A' in explicit recognition studies. It is becoming increasingly common for researchers to employ A' and journal editors and reviewers to suggest that data be expressed in terms of A' , under the belief that the metric requires fewer assumptions about the origin of the data than do other measures, such as d' . Macmillan and Creelman (1996) have recently provided evidence demonstrating that this belief is mistaken. In addition, the fact that A' systematically deviates across subjective reports in a manner predicted by dual-process theory demonstrates that the implied ROC of the metric is not consistent with actual empirical data (see Swets, 1986, for discussion). This suggests that its use should be expressly avoided as a summary of explicit recognition across conditions or subject groups, in which remembering/recollection may differ (dual-process) or for which different response criteria may be adopted (unequal variance signal detection). Importantly, the present findings and conclusions are not confined to A' but, in principle, apply to *any* single-point performance measure that implies an ROC with a rigid surface. Aside from A' , the three most popular single-point indices are the double equal-threshold (DET) correction (i.e., hits – false alarms), the single high-threshold (SHT) correction [i.e., (hits – false alarms) / (1 – false alarms)], and the equal variance signal detection d' . The DET mea-

sure assumes a rigid linear ROC with a slope of one, which like A' , moves upward along the minor diagonal as performance increases (see Swets, 1986, for details). The SHT measure is also rigid but intersects the upper right corner of the ROC plot. As performance increases, this measure assumes a rigid surface that tilts upward and is "hinged" in the upper right corner. Similar to A' , these single-point estimates are rigid, and therefore, one would predict that the discrepancy between the correction for overall recognition and that for remembering with use of these measures should also systematically change as a function of the remember rate and overall performance level in the same manner as was predicted for A' . The last rows of Table 2 confirm this prediction for these measures. As was the case with A' , there is a negative relation between the remember rate and the single-point measure discrepancy and a positive relation between overall performance and the discrepancy. Given this, these single-point corrections should also be avoided as a summary of recognition performance across conditions or groups, for the same reason as A' . Importantly, the critical property of both A' and the other single-point corrections that makes them poor candidates as summaries of recognition makes them very useful for evaluating more complex models of memory—namely, they embody a rigid and well-defined response surface that can serve as a benchmark for determining how the empirical ROC is systematically deviating from this surface as a function of different response types, levels of confidence, or other manipulations. In conjunction with the statistical control afforded by multiple regression, this approach may be useful in augmenting the traditional fitting techniques typically used in ROC analysis.

Finally, these findings suggest that it may be unlikely that overall recognition performance can be effectively summarized with any single index, because a full characterization of the ROC requires that one be able to determine at least two characteristics: (1) the area below the response surface and (2) the asymmetry of that surface. Currently, these two characteristics are typically captured by using either an unequal variance signal detection approach or a dual-process approach to old/new recognition, both of which rely on *two separate factors* to effectively model the asymmetric shape of the ROC and explain performance. This is made somewhat more explicit in the dual-process approach (viz., recollection and familiarity) but is nevertheless also true in the unequal variance model, because the variance of the old-item distribution (σ^2) is usually allowed to freely change regardless of the distance between the old- and the new-item distributions (i.e., d') when fitting the model to empirical data. This is not to say that two-factor models cannot be reduced to a single value for pragmatic reasons. The signal detection measures d'_e and A_z (see Macmillan & Creelman, 1991) represent such cases. However, by definition, these measures result in loss of information, because they are derived from estimates of both the distance between distributions and the variance ratio of those distributions. Given the two estimates, one can compute the accuracy value; however,

the reverse is not true. Two equivalent values of A_z or d'_e do not demand or even imply equivalent underlying distances or variance ratios. Thus, both dual-process and unequal variance approaches would contend that there is a critical loss of information whenever performance is summarized with a single index, because the relative contribution of two factors is necessarily conflated within the single value. Regardless of whether one advocates a dual-process or an unequal variance signal detection approach, the present data demonstrate that the observed remember rate of each individual provides important information regarding underlying memory processes that could not be gleaned from the overall hit and false alarm rates in isolation.

REFERENCES

- DOBBINS, I. G., KHOE, W., YONELINAS, A. P., & KROLL, N. E. A. (2000). Predicting individual false alarm rates and signal detection theory: A role for remembering. *Memory & Cognition*, **28**, 1347-1356.
- DONALDSON, W. (1996). The role of decision processes in remembering and knowing. *Memory & Cognition*, **24**, 523-533.
- DUZEL, E., YONELINAS, A. P., MANGUN, G. R., HEINZE, H. J., & TULVING, E. (1997). Event-related brain potential correlates of two states of conscious awareness in memory. *Proceedings of the National Academy of Science*, **94**, 5973-5978.
- GARDINER, J. M., & GREGG, V. H. (1997). Recognition memory with little or no remembering: Implications for a detection model. *Psychonomic Bulletin & Review*, **4**, 474-479.
- GRIER, J. B. (1971). Nonparametric indexes for sensitivity and bias: Computing formulas. *Psychological Bulletin*, **75**, 424-429.
- HENSON, R. N. A., RUGG, M. D., SHALLICE, T., JOSEPHS, O., & DOLAN, R. J. (1999). Recollection and familiarity in recognition memory: An event-related functional magnetic resonance imaging study. *Journal of Neuroscience*, **19**, 3962-3972.
- HINTZMAN, D. L. (1994). On explaining the mirror effect. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **20**, 201-205.
- HIRSHMAN, E., & HENZLER, A. (1998). The role of decision processes in conscious recollection. *Psychological Science*, **9**, 61-65.
- HIRSHMAN, E., & MASTER, S. (1997). Modeling the conscious correlates of recognition memory: Reflections on the remember-know paradigm. *Memory & Cognition*, **25**, 345-351.
- INOUE, C., & BELLEZZA, F. S. (1998). The detection model of recognition using know and remember judgments. *Memory & Cognition*, **26**, 299-308.
- KHOE, W., KROLL, N. E. A., YONELINAS, A. P., DOBBINS, I. G., & KNIGHT, R. T. (2000). The contribution of recollection and familiarity to yes-no and forced-choice recognition tests in healthy subjects and amnesics. *Neuropsychologia*, **38**, 1333-1341.
- KUČERA, H., & FRANCIS, W. N. (1967). *Computational analysis of present-day American English*. Providence, RI: Brown University Press.
- MACMILLAN, N. A., & CREELMAN, C. D. (1991). *Detection theory: A user's guide*. New York: Cambridge University Press.
- MACMILLAN, N. A., & CREELMAN, C. D. (1996). Triangles in ROC space: History and theory of "nonparametric" measures of sensitivity and response bias. *Psychonomic Bulletin & Review*, **3**, 164-170.
- MANDLER, G. (1980). Recognizing: The judgment of previous occurrence. *Psychological Review*, **87**, 252-271.
- PEDHAZUR, E. J. (1997). *Multiple regression in behavioral research*. Orlando: Harcourt Brace.
- POLLOCK, I., & NORMAN, D. A. (1964). A non-parametric analysis of recognition experiments. *Psychonomic Science*, **1**, 125-126.
- RAJARAM, S., & ROEDIGER, H. L., III. (1997). Remembering and knowing as states of consciousness during retrieval. In J. D. Cohen & J. W. Schooler (Eds.), *Scientific approaches to consciousness* (pp. 213-240). Hillsdale, NJ: Erlbaum.
- RATCLIFF, R., SHEU, C., & GRONLUND, S. D. (1992). Testing global memory model using ROC curves. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **99**, 518-535.

- STRETCH, V., & WIXTED, J. T. (1998). Decision rules for recognition memory confidence judgments. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **24**, 1397-1410.
- SWETS, J. A. (1986). Indices of discrimination or diagnostic accuracy: Their ROCs and implied models. *Psychological Bulletin*, **99**, 100-117.
- TULVING, E. (1985). Memory and consciousness. *Canadian Psychology*, **26**, 1-12.
- 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., DOBBINS, I. G., SZYMANSKI, M. D., DHALIWAL, H. S., & KING, L. (1996). Signal detection, threshold, and dual-process models of recognition memory: ROCs and conscious recollection. *Consciousness & Cognition*, **5**, 418-441.

NOTES

1. For the development of the "nonparametric" A' measure of accuracy, see Pollack and Norman (1964) or Grier (1971). Macmillan and Creelman (1996) recently demonstrated that A' , as with other measures of accuracy, does in fact require assumptions regarding the underlying distributions that govern performance and, therefore, is not adequately described as "nonparametric." The computational formula for the metric when performance is above chance is

$$A' = .5 + .25 \left[\frac{(H - F)(1 + H - F)}{H(1 - F)} \right],$$

where H is the hit rate and F is the false alarm rate.

2. The invariance in the shape of the A' ROC will necessarily break down as performance approaches chance and the ROC is forced to "flatten out" along the major diagonal.

3. For simplicity, I have assumed recollection to be faultless during the graphical discussion of the predictions. However, because the actual data contain incorrect remember responses, incorrect remember responses are

also explicitly modeled in the simulations. Because incorrect remember responses mirror the introspective and electrophysiological properties of correct remembering (Duzel et al., 1997; Rajaram & Roediger, 1997) and because, despite their faulty nature, they involve the retrieval of actual prior contextual information, I have chosen to model them as separate from familiarity-driven errors and as arising from the same process that enables correct remembering.

4. It is not necessary that the independent variable for overall performance in the regression equation be characterized with the same accuracy measure as the dependent variable. Any measure of overall performance that increases reliably as a function of the total area under the ROC will yield similar results and conclusions. For example, the perpendicular distance from the overall performance location to the chance performance line,

$$D = \frac{\sqrt{2}H_i}{2},$$

where H_i is the difference between the overall hit and false alarm rate (i.e., hits — false alarms), results in a very similar pattern of regression results when used in place of A' for overall recognition as a predictor. Identical regression results are obtained if one uses the double equal-threshold (DET) correction (hits — false alarms) as the independent variable of overall performance, because the distance measure D above is a strictly increasing function of the DET correction value. Finally, similar results are also observed if one uses the equal variance d' correction as an overall performance indicator.

5. The equations used for calculating the likelihoods and criterion locations were taken from Stretch and Wixted (1998; see the original for development):

$$L = re^{-0.5r^2(f - d')^2 + 0.5f^2},$$

where $r = 1/s$, the inverse of the variance ratio of old and new items distributions and

$$f = \frac{d'r^2 \pm \sqrt{(d'r^2)^2 - (r^2 - 1)[(d'r)^2 - 2\log(r/L)]}}{(r^2 - 1)}.$$

APPENDIX

Experiment 1

These data were previously published in Yonelinas et al. (1996). Subjects were presented with 160 medium- to high-frequency words drawn from the Toronto word pool and were instructed to try to remember the gender of the speaker for each, half of which were presented in a male voice and the other half in a female voice. In addition, male-voiced items were presented twice. The effects of this manipulation is not of interest, and the data are collapsed across conditions. During testing, the subjects were presented the initial 160 items along with an additional 80 new items in random order. For each item they made a 6-point confidence judgment similar to that in Experiment 1, followed by a remember-know judgment for items judged *old*. The subjects were instructed that remembering need not pertain only to the gender of the study list speaker, but could also apply to any other retrieved aspect of the study event.

Experiment 2

These data are from a study examining differences between yes/no and forced-choice responding (Khoe, Kroll, Yonelinas, Dobbins, & Knight, 2000). The subjects were presented with 200 six-to-eight letter words and were asked to generate two vowels not present in each word. Words were presented one at a time, and the subjects were instructed to proceed at their own pace. After the study phase, the subjects engaged in a spatial distractor task for 30 min. Following the retention interval, the subjects were tested for their recognition memory in blocks of 200

yes-no and 100 two-alternative forced-choice (2AFC) trials. Only the data from the yes-no test are included here. The blocks of yes-no and the 2AFC blocks were counterbalanced across subjects. In the yes-no condition, 100 studied words were randomly mixed together with 100 *new* distractors. One word was presented on each trial, and the subjects were asked to decide whether the word was an *old* or a *new* word. Following each *old* judgment, they were asked to make a remember/know judgment.

Experiment 3

Subjects were presented with 150 low-frequency words ($M = 8.65$, Kučera & Francis, 1967), during which they indicated via computer mouse how many syllables each item had. Study rate was self-paced, with a new item appearing after each response. After the final word of the study list, the subjects were given a recognition test in which half the items were old and the other half new (total, 300). For each item, the subjects were required to make a recognition confidence judgment from 1 (*certain old*) to 6 (*certain new*). In addition, following each *old* response, they were asked to decide whether they could remember anything specific about the item's previous occurrence—for example, anything they thought about or noticed during the previous presentation, including making a specific overt syllable judgment. The data analyzed in the present study are collapsed across response confidence.