

Predicting individual false alarm rates and signal detection theory: A role for remembering

IAN G. DOBBINS, WAYNE KHOE, ANDREW P. YONELINAS, and NEAL E. A. KROLL
University of California, Davis, California

The relationships between hit, remember, and false alarm rates were examined across individual subjects in three remember-know experiments in order to determine whether signal detection theory would be consistent with the observed data. The experimental data differed from signal detection predictions in two critical ways. First, remember reports were unrelated, or slightly negatively related, to the commission of false alarms. Second, both response types (rememberers and false alarms) were uniquely related to hit rates, which demonstrated that the hit rate cannot be viewed as the result of a single underlying strength process. These results are consistent with the dual-process signal detection model of Yonelinas (1994), in which performance is determined by two independent processes—retrieval of categorical context information (remembering) and discriminations based on continuous item strength. Remember and false alarm rates selectively tap these processes, whereas the hit rate is jointly determined. Monte Carlo simulations in which the dual-process model was used successfully reproduced the pattern in the experimental data, whereas simulations in which a signal detection model, with separate “old” and “remember” criteria, was used, did not. The results demonstrate the utility of examining individual differences in response types when one is evaluating memory models.

Typically, memory researchers use the theory of signal detection in an attempt to remove the unwanted effects of response bias in comparing accuracy across experimental conditions or different subject groups (for a review, see Macmillan & Creelman, 1991). In such instances, it is not the theory itself which is of interest, but rather the effects of encoding manipulations or group differences (e.g., control subjects vs. patients with neurological damage) on a theory-based measure of accuracy— d' . Another recent approach has been to use the theory itself as an explanation of dissociations in response types across different remember-know studies (e.g., Donaldson, 1996). The remember-know paradigm, originally developed by Tulving (1985), requires subjects not only to select which items they believe are “old” during recognition testing, but also to select one of two responses regarding the conscious basis of why they believe a particular item is in fact old. In the case of “remember” responses, subjects are assumed to access recollective information regarding previous thoughts, feelings, or perceived events that have occurred during their earlier encounter with the item. If they are unable to retrieve such information but have a strong sense that the item was previously studied, they are instructed to respond “know.” A common example of “knowing” often provided to subjects is the phenomenon of seeing someone who strikes one as highly familiar, but being

unable to identify the contextual basis or source of this familiarity (Mandler, 1980).

Remember-know studies have led to a myriad of dissociations between remembering and knowing responses (for a review, see Rajaram & Roediger, 1996). Such dissociations are consistent with the assertion that remembering and knowing reflect different cognitive processes or states of consciousness. However, Donaldson (1996) presented a slight modification of signal detection theory that could accommodate dissociations of remembering and knowing reports without postulating different underlying memory processes. This model, shown in Figure 1, will be referred to as the two-criterion signal detection model.

The model is a standard signal detection model with two strength criteria: one for determining an item as old (C_o) and another more stringent criterion (C_r) for providing remember responses. More specifically, “items that lie above that criterion [C_o] are identified as having previously occurred (i.e., are given a yes response). A second criterion is then established that divides the yes responses into those above the new criterion, which get labeled ‘remember,’ and those below it, which get labeled ‘know.’” (Donaldson, 1996, p. 524). In addition to the two criteria, separate groups of subjects (each with a similar overall accuracy) can be modeled as more or less conservative in their responses. The top, middle, and bottom panels of Figure 1 represent conservative, neutral, and liberal groups of subjects, respectively. As one goes from the most to the least conservative groups in this particular example, it can be seen that the correct remember rate increases, whereas the correct know rate decreases. In contrast, both incorrect remember and know responses increase

This research was supported in part by National Institutes of Health Grants MH59352-01 to A.P.Y. and NS17778 to N.E.A.K. Address correspondence to I. G. Dobbins, MGH-NMR Center, Department of Radiology, Bldg. 149, 13th St., Mail Code 149 (2301), Charlestown, MA 02129-2060 (e-mail: ian@nmr.mgh.harvard.edu).

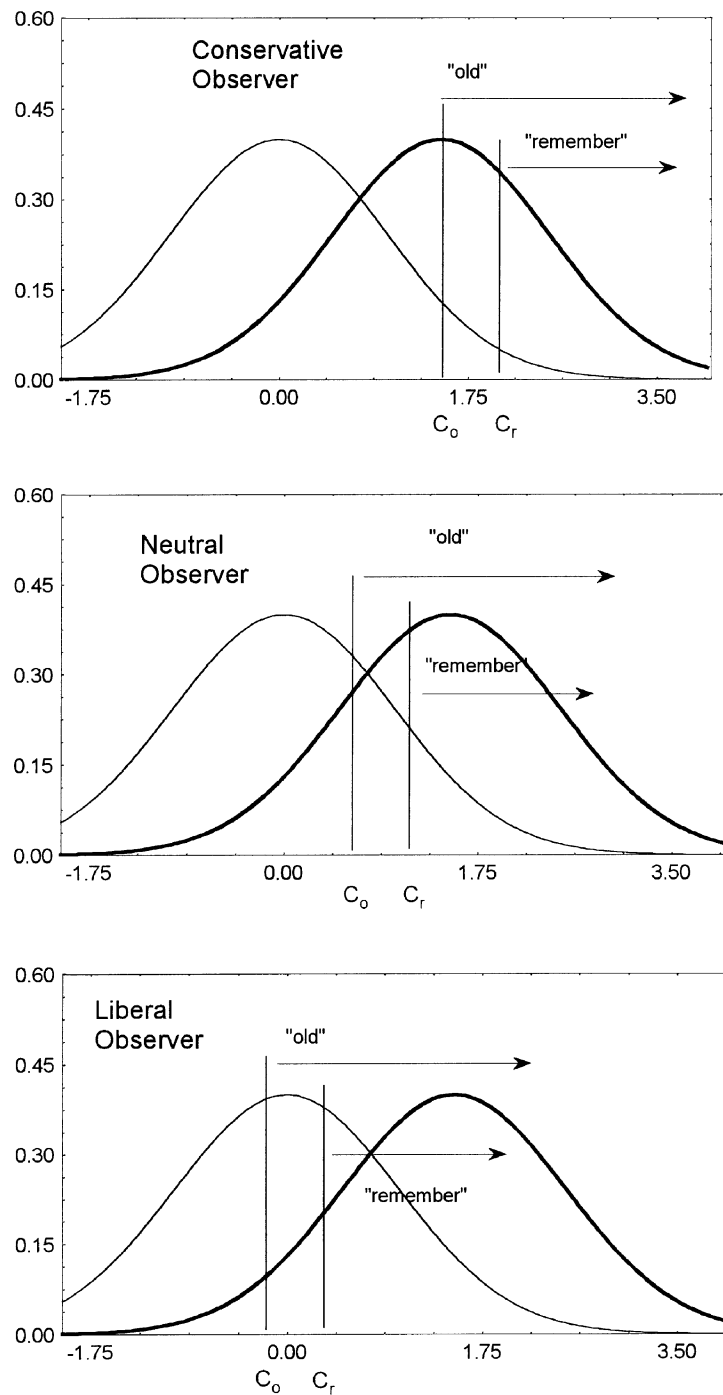


Figure 1. The signal detection model with old/new and remember criteria. Top, middle, and bottom panels represent conservative, neutral, and liberal observers/groups, respectively.

across the groups. Thus, the model is quite flexible, and it has had considerable success in describing the outcomes of a number of remember-know studies (e.g., Hirshman & Henzler, 1998; Hirshman & Master, 1997; Inoue & Bellezza, 1998; for exceptions, see Gardiner & Gregg, 1997).

However, if correct, the two-criterion signal detection model should not be restricted to the description of group

level differences because it also makes testable predictions regarding the possible patterns of hits, false alarms, and remember rates that should be observed across individual subjects. That is, the three panels of Figure 1 can be thought of as representing a conservative, a neutral, and a liberal individual within a particular experiment, each with a similar accuracy that is normally distributed

around a central d' score. If we assume that there is noticeable variability in where each subject places his or her old/new criterion within a given experiment, the model in Figure 1 predicts a positive relationship between subjects' hit and false alarm rates. Liberal responders will have both high hit and false alarm rates, and the reverse would be true for more conservative subjects.¹

More interesting predictions can be made regarding remember rates.² For example, the constraint that each subject must place his or her remember criterion higher on the strength continuum than the old/new criterion will mean that on average the two criteria (and hence the proportion of correct old and remember response types) will be correlated across subjects. Thus conservative observers will not only tend to commit fewer false alarms, but will also be less likely to report remembering; the reverse would be true for liberal observers. Importantly, the relation between remember and false alarm rates would be entirely indirect and merely the result of the criterion placement restriction. That is, because subjects adopt their remember criterion as a function of their old/new criterion location, the former has no direct bearing on the observed false alarm rate. Therefore, any information regarding false alarms present in the remember rate is redundant to that present in the hit rate.

In summary, visual inspection of the two-criterion signal detection model illustrated in Figure 1 suggests three clear predictions regarding the relationships between hit, remember, and false alarm rates for individual subjects. First, hit and false alarm rates will be positively correlated via the old/new response criterion. Second, remember and false alarm rates will be positively correlated because of the relationship between the remember and old/new response criteria. Finally, because the relation between remember and false alarm rates is entirely redundant to that between the hit and false alarm rates, one would expect multiple regression to demonstrate that correct remember rates will yield no unique predictive information about false alarm rates. This last prediction of complete redundancy will be more fully expanded in the Results section.

In the remainder of the paper, we will examine the individual response data from three separate remember-know studies, more formally examine the predictions of the two criterion signal detection model using Monte Carlo generated remember-know data, and discuss the implications of the results for both the two-criterion signal detection model and the false alarm behavior of individual subjects during remember-know recognition paradigms. First we will report on a consistent pattern of response relationships observed in the individual subject data of three separate remember-know experiments.

METHOD

Experimental Data

The following data are drawn from three unrelated studies in which the primary focus was to compare group accuracy differences,

or the estimation of recollection processes, and not to analyze individual responses. They are similar in that each contained 24 subjects and was administered via computer with keyboard input by the subjects. A brief description of each experiment is given prior to the descriptions of results of the correlation and regression analyses. Significant effects are reported at the .05 level unless otherwise noted. Figure 2A shows the scatterplots relating hit and false alarm rates, hit and remember rates, and remember and false alarm rates.

Experiment 1. These data were previously published in Yonelinas, Dobbins, Szymanski, Dhaliwal, and King (1996). The subjects were presented with 160 medium- to high-frequency words drawn from the Toronto word pool; they were instructed to try to remember the gender of the speaker for each word. Half of the words were presented in a male voice, the other half in a female voice. In addition, male-voiced items were presented twice. The effects of this manipulation are not of interest and the data are collapsed across conditions. During testing, the subjects were presented the initial 160 items with an additional 80 new items in random order. Items were counterbalanced by constructing three lists of 80 items and rotating these across conditions (male, female, and new) for each subject. For each item, the subjects made a six-point confidence judgment, followed by a remember-know judgment for items judged as being old. The subjects were instructed that their remembering need not only pertain to the gender of the 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. The 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 2AFC trials were counterbalanced across subjects. The items were counterbalanced by constructing four lists of 200 items from an 800-item word pool, each of which was rotated through possible study test permutations 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 old or new. Following each *old* judgment, they were asked to make a remember-know judgment.

Experiment 3. The subjects were presented with 150 low-frequency words ($M = 8.65$; Kučera & Francis, 1967) during which time they indicated via computer mouse how many syllables each item had. Items were counterbalanced using six randomized study/test lists and two randomized test orders. 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 (1 = *certain old* to 6 = *certain new*). In addition, following each *old* response, they were asked to decide whether they remembered 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 were collapsed across response confidence.

Results

Table 1 shows the correlation and regression data for each of the three experiments; Table 2 shows the means and standard deviations. In all three experiments, there

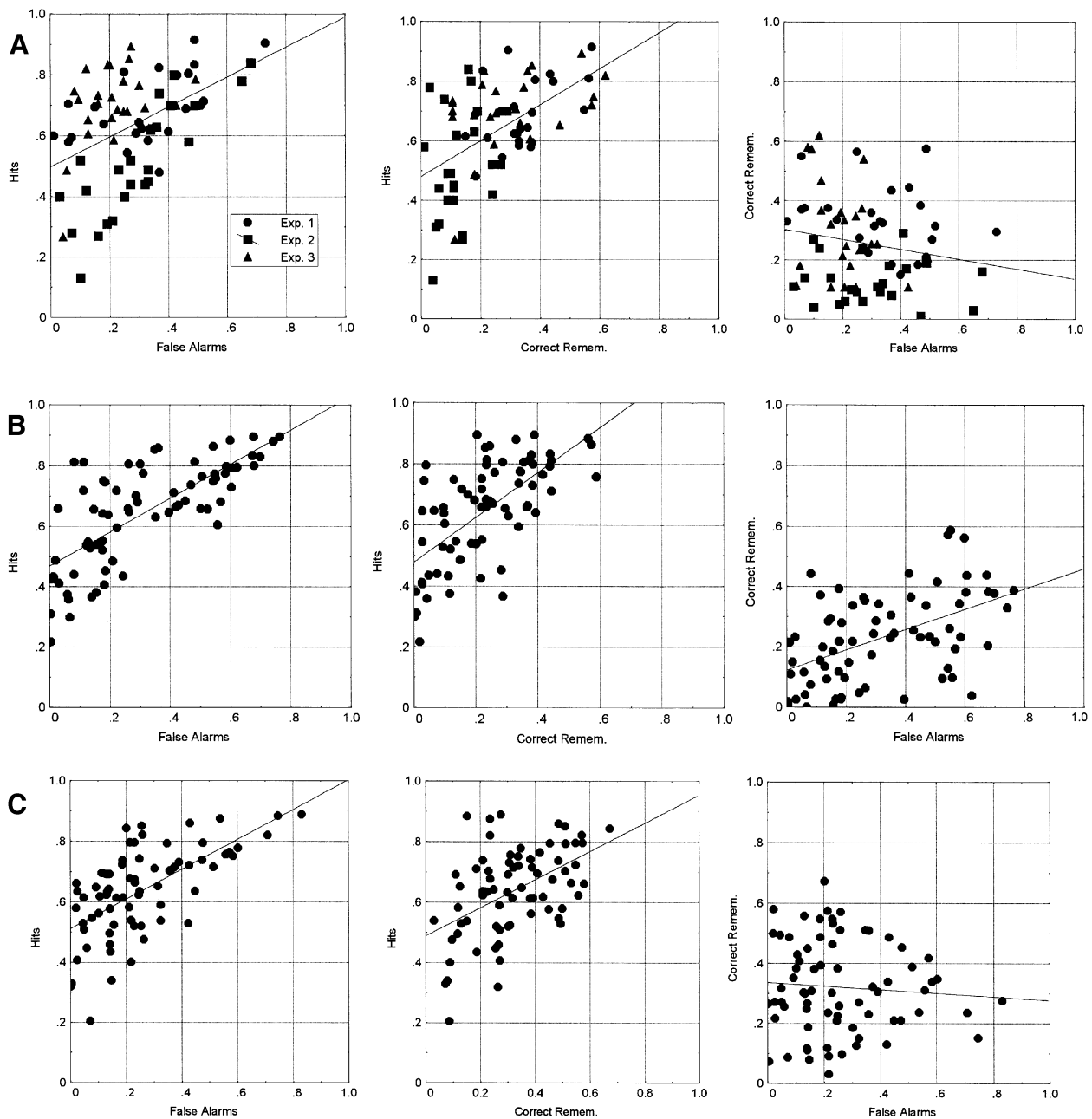


Figure 2. Scatter plots of experimental and Monte Carlo data. The left plots show the relation between hit and false alarm rates, the middle between hits and correct remember rates, and the right between correct remember and false alarm rates. The top panels show the experimental data, the middle the Monte Carlo data for the two-criterion signal detection model, and the bottom the Monte Carlo data for the dual-process model with a false remember rate of 0. The lines are best fit linear regressions.

was a significant positive correlation between the hit and false alarm rates, which indicates that the adopted old/new criterion varied extensively across subjects. Remember and hit rates were also positively correlated. In Experiments 1 and 3, this correlation was significant, whereas in Experiment 2, the correlation was positive, although the obtained *p* value was only .20. Finally, in none of the three experiments was there any evidence for a positive relationship between the remember and false alarm rates

since all three showed negative, but nonsignificant, correlations. These correlations are displayed in scatterplots for the aggregate data in Figure 2A.

The two-criterion signal detection model accounts for these findings only partially. The fact that hits and false alarms were positively related is consistent with the signal detection assumption of a single old/new response criterion, which determines both, and which varies across each subject. The positive relation between remember and hit

Table 1
Regression Results

	Hits		Remember		ΔR^2	Correlations		
	β	<i>SE</i>	β	<i>SE</i>		H/FA	H/Rem	Rem/FA
Experiment 1	.88*	.14	-.69*	.14	.36*	.56*	.47*	-.26
Experiment 2	.91*	.10	-.33*	.10	.10*	.83*	.27	-.08
Experiment 3	.68*	.17	-.58*	.17	.29*	.44*	.40*	-.31
Total exp. data	.79*	.10	-.60*	.10	.26*	.46*	.54*	-.18
2-Criterion SDT sim. ^a	.70*	.11	.04	.11	.001	.73*	.65*	.50*
Dual-process sim. (.00) ^b	.88*	.09	-.50*	.09	.19*	.63*	.49*	-.07
Dual-process sim. (.15)	.83*	.08	-.24*	.08	.05*	.74*	.40*	.09
Dual-process sim. (.30)	.63*	.10	.08	.10	.005	.65*	.40*	.33*

Note— ΔR^2 is the gain in R^2 when correct remember rate is entered into the regression equation. ^aTwo-criterion signal detection model. ^bThe parentheses indicate the three levels of false remember rates modeled in dual process simulations. Remember rate is for correct responses only. * $p < .05$.

rates would then result from the theory constraint that the remember criterion be placed to the right of the old/new criterion. Thus, on average, both criteria in Figure 1 will “track” together across subjects. The subjects who adopted a liberal old/new criterion would be more likely to adopt a liberal remember criterion; the reverse would be true for the conservative subjects. However, if this was the case, one would also expect to see a positive (although indirect) relation between remember and false alarm rates. If remembering is related to the hit rate, and the hit rate is related to the false alarm rate, the remember rate should be related to the false alarm rate. There are two possible arguments against this line of reasoning. The first is that there simply may not be enough variability in the remember rates to detect a relationship. For example, if the subjects had adopted the remembering criterion to the far right in Figure 1, there would be insufficient remember rates for detecting a correlation. A second concern is that the remember rates might be determined in a region of the old-item distribution that is severely nonlinear in comparison with the portion of the new-item distribution determining the false alarm rates. We will directly address these possible objections using the outcomes of Monte Carlo simulations presented later.

The second analysis that we performed involved multiple regression. As discussed previously, the two-criterion signal detection model suggests that remember rates would be largely inconsequential in one’s predicting sub-

jects’ false alarm rates because the rates are determined secondarily to the old/new decision process. Because multiple regression determines the unique contribution of the predictor variables, one’s entering remember rates into the regression equation relating hits and false alarms should yield no improvement because the information in the remember rate is redundant to what is already present in the hit rate. This null effect would seem even more likely in the present data set, because the correlation between remember and false alarm rates in the raw data is not significantly different from zero. However, when the remember rate was entered into the regression for each experiment, three important and related effects were observed. First, in each case that the variance was accounted for by the regression equation, R^2 was sizably increased (see Table 1). Second, the apparent relationship between the hit and false alarm rates increased. This can be seen by comparing the beta weights for the hit rates with the zero order correlations between hits and false alarms. In each case the beta weight is considerably higher than the correlation in the raw data. Finally, the beta weights for remember rates are negative *and* significant, despite the fact that there was no significant relation between remember and false alarm rates evident in the raw data. This pattern is also observed in the aggregate data.

In the multiple regression literature, this pattern of effects is known as *traditional suppression*, and the remember rate is labeled a *suppressor variable* (for reviews, see

Table 2
Means and Standard Deviations for Experiments and Simulations

	Hits		FAs		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>
	Experiment 1	.69	.12	.33	.18	.34	.11	.05
Experiment 2	.51	.19	.30	.17	.13	.08	.05	.05
Experiment 3	.70	.13	.21	.11	.30	.16	.03	.03
Total exp. data	.63	.17	.28	.16	.26	.15	.04	.05
2-Criterion SDT sim.	.65	.17	.32	.22	.23	.14	.07	.08
Dual-process sim. (.00) ^a	.64	.15	.26	.19	.32	.15	n/a	n/a
Dual-process sim. (.15)	.65	.14	.34	.20	.26	.11	.05	.02
Dual-process sim. (.30)	.58	.14	.33	.16	.21	.10	.09	.04

Note—Parentheses in column 1 indicate false remember rates for each simulation.

Conger, 1974; Pedhazur, 1997; Wiggins, 1973). Suppression occurred in the present data because the remember rate shared a relationship with the hit rate that was independent of false alarm behavior. In short, the subjects with high remember rates can be expected to have high hit rates, but this increase in remember rates does not contribute to the tendency to commit false alarms. The role of the remember rate then is to literally subtract the variance in the hit rates that is not predictive of false alarms—hence the negative beta weight and the term *suppression*. This effect can be demonstrated graphically with Venn diagrams (see Figure 3). The top panel shows the most common case in regression, partial redundancy, in which the predictors (hit and remember rates) are themselves correlated. When the unique contribution of each is calculated via regression (shaded areas in Figure 3), it is shown to be smaller than that apparent in the raw data. In contrast, the middle panel shows the complete redundancy predicted by the two-criterion signal detection model. Because the remember rate contains no unique information in regard to the relation between hits and false alarms, its predicted regression weight is zero. Finally, the present suppression effect is demonstrated in the bottom panel. Although hit rates are correlated with remember rates, this relationship does not overlap with that between hit and false alarm rates.

The suppression effect of remember rates poses a problem for the signal detection account because it demonstrates that the hit rate has at least *two* unique and systematic sources of variance—one resulting from the relationship between remember and hit rates, the other between the remaining portion of hit rates (after remember rate is subtracted) and false alarm rates. This suggests that the observed hit rate cannot be viewed as the result of a single underlying strength process. Next, we will demonstrate that the suppression effect cannot be the result of nonlinearities that might arise from a single underlying signal detection process.

The following Monte Carlo simulations were conducted to determine whether the two-criterion signal detection model would generate remember-know data that would be consistent with the four key empirical findings outlined previously. These findings consist of (1) a positive correlation between hit and false alarm rates, (2) a positive correlation between remember and hit rates, (3) a null (or slightly negative) correlation between remember and false alarm rates, and most importantly, (4) the remember rate suppression effect. We were unable to find a remember criterion placement that was consistent with the account of Donaldson (1996) and that was capable of mirroring the present effects—in particular, the null or slightly negative remember/false alarm correlation and

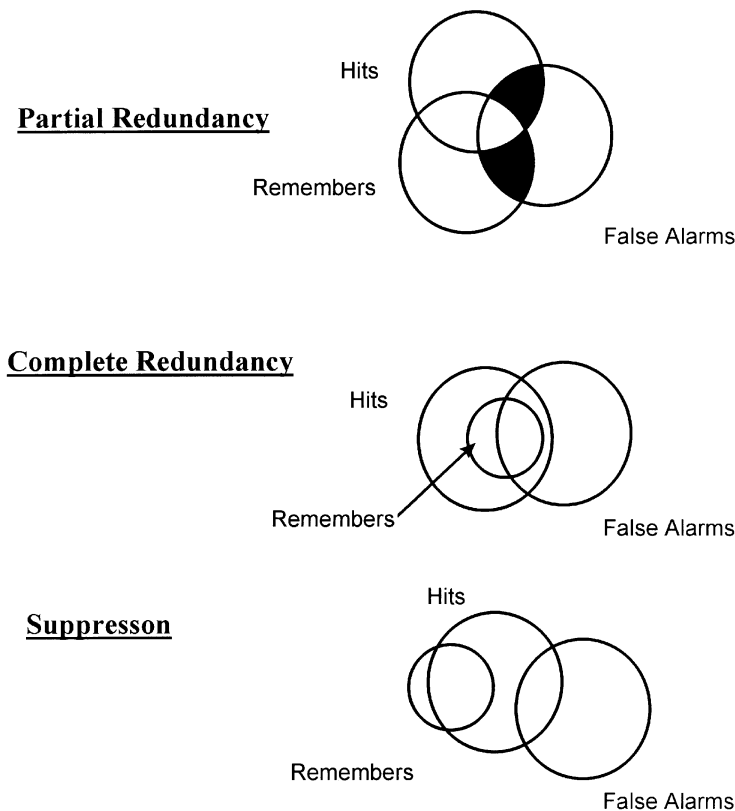


Figure 3. Venn diagrams of partial redundancy, complete redundancy, and suppression.

the remember rate suppression effect. In contrast, when we modeled the hit rate as being the result of independent recollection (binomial) and familiarity (signal detection) processes, and the false alarm rate as being largely determined by the familiarity process, the generated data were hard to distinguish from the experimental results.

Monte Carlo Results

Two-criterion signal detection model. Within the theory of signal detection there is a straightforward, though intrinsically nonlinear, relationship between a given subject's hit and false alarm rates. Both rates are determined by the observer's decision criterion, labeled C_o in Figure 1. During testing, old and new items that are associated with familiarity or memory strength values above this cutoff will elicit positive responses.

To simulate this model, we assumed that subjects with similar memory abilities (say freshmen in an introductory psychology course) who participate in similar encoding conditions will have a d' that is normally distributed around a central value.³ In order to closely match the average experimental data, each of the 72 artificial subjects was given a d' that was sampled from a normal distribution with a mean of .95 and a standard deviation of .50.⁴ In order to mirror the frequency and shape of hit rates in the subject data, an old/new response criterion was determined by randomly choosing from a normal distribution of hit rate values with a mean of .70 and a standard deviation of .20. The sampling process was repeated for hit rates that fell below .13 or above .92 (the actual range in the data). This yielded a slightly left-skewed distribution of hit rates with a mean, shape, and standard deviation similar to those of the actual data and determined the location of the old/new response criterion.

In order to obtain remember rates that were consistent with the description by Donaldson (1996), in which subjects first chose the old/new criterion and then selected a more conservative criterion for remember reports, the remember criterion (C_r) was sampled to fall approximately 1.25 standard deviations above the old/new criterion (drawn from a normal distribution with a standard deviation of .50); values falling below the old/new criterion due to chance were resampled in order to meet the theory constraint that the remember criterion must fall to the right of the old/new criterion for each subject.⁵ This selection process resulted in remember rates similar in mean and standard deviation to those observed in the actual data. By using these parameters, we then calculated each subject's false alarm rate on the basis of the location of the old/new criterion. The data for a typical run are shown as scatterplots in Figure 2B, and the appropriate values are listed in Tables 1 and 2. Despite the fact that the underlying signal detection process is intrinsically nonlinear, all three panels of Figure 2B show significant linear relationships. Importantly, the right panel of Figure 2B shows a sizable linear relation ($r = .50, p < .001$) between remember and false alarm rates across subjects and, therefore, the lack of a positive correlation between

remember and false alarm rates in the actual data is inconsistent with signal detection theory, given the level of observed remember rates. The second important result of the simulation can be seen when both the hit and remember rates are entered into a multiple regression (see Table 1). By construction, the remember rate has no actual bearing on the observed false alarm rate, which is determined by the previously established old/new response criterion, C_o . In other words, the observed raw correlation between remember and false alarm rates is an artifact of the constraint that the remember criterion lies above the old/new criterion. This was borne out in the results of the multiple regression. Examination of the regression coefficients showed that all predictive value lay with the hit rates; the coefficient for remember rate was not significantly different from zero and yielded no unique contribution to predicting false alarms. This pattern differed from the experimental data in which inclusion of the remember rate yielded a significantly negative beta weight, significantly increased R^2 , and elevated the apparent relationship between hits and false alarms.

The previous simulation was based on the assumption that the old- and new-item familiarity distributions have an equal variance. This need not be the case, and signal detection theorists often assume that the old-item distribution is in fact more variable than the new (e.g., Ratcliff, Sheu, & Gronlund, 1992). This was explicitly modeled by repeating the simulation with standard deviations for the old-item distribution between 1 and 1.5 times that of the new. The results were essentially unchanged. Remember and false alarm rates remained correlated in the raw scores, and the regressions demonstrated redundancy, not suppression.

The dual-process model. The previous Monte Carlo simulations have demonstrated that the signal detection model by itself fails to adequately account for the pattern of relationships observed in the experimental data. The purpose of the next simulation was to see whether the dual-process signal detection model of Yonelinas (1994) could reproduce the present experimental results. The model assumes independent, parallel contributions of recollection and familiarity processes to the observed hit rates of each subject. Recollection involves the retrieval of contextual information associatively linked with the item during a previous study episode. For example, when seeing an item at test, a subject may recollect his/her initial mispronunciation of the item during study. Because recollection is assumed to be a threshold process, it can be measured as probability. It is this type of information that subjects are presumed to be accessing when they give remember reports. In contrast, familiarity is assumed to operate in accordance with the equal-variance signal detection model presented in Figure 1 (without the need for a second remember criterion).

The Monte Carlo process was similar to the previous signal detection runs with some exceptions. The starting d' was now sampled from a population centered on .53 with a standard deviation of .50. The reason for this re-

duction (d' was .95 in the previous simulation) was that the observed hit rate was assumed to reflect the independent contribution of *both* familiarity and recollection processes [i.e., $\text{hits} = R + F - (R * F)$]. Therefore, the observed hit rate was an overestimate of the true contribution of signal detection to successful responding. In order to estimate the actual proportion of hits due to familiarity, we assumed a recollection rate of .30 (on the basis of the observed total remember rate) and solved the equation for F yielding .47. This value was then used in conjunction with the observed false alarm rate of .28 so that we could determine an approximate d' starting value of .53 to center the sampling process.

The old/new signal detection criterion was fixed by sampling hit rates between .01 and .99 that were normally distributed with a mean of .5 and standard deviation of .2. A separate and independent remember rate was sampled from a population centered on the observed mean remember rate of .30 and a standard deviation of .17. In addition, remember rates were constrained to fall between the actual observed range of .02 to .71. Following this, the subject's total hit rate was determined by independently combining the sampled remember rate and signal detection hit rate [i.e., $R + F - (R * F)$]. If the total hit rate fell above or below the observed range in the actual data (.13–.92), the process was repeated. The false alarm rate was then determined by calculating the portion of the new-item distribution that fell above the sampled old/new signal detection criterion.

The data for a run are shown as scatterplots in Figure 2c, and the appropriate values are listed in Tables 1 and 2. The correlations between the hit, false alarm, and remember rates are statistically indistinguishable from those observed in the actual experimental data. More importantly, the regression results also mirror the observed data. When the remember rate was included in the regression equation, a clear suppression effect was observed. The proportion of variance accounted for (R^2) increased, the apparent relation between hit and false alarm rates increased, and the regression weight for remember rates was significant and negative, despite the fact that the correlation in the raw data was not significantly different from zero (see Table 1).

The preceding simulation was based on the assumption that recollection solely contributes to the hit rate. However, the actual data show a small proportion of incorrect remember responses, and experimental data continue to accumulate, suggesting that in some instances both recollective and familiarity processes can contribute to incorrect responding (e.g., Duzel, Yonelinas, Mangun, Heinze, & Tulving, 1997; Roediger, McDermott, & Robison, 1998). In order to model this, we assumed that some proportion of each subject's sampled remember rate was faulty. This proportion was then subtracted from the remember rate to determine the correct remember rate and was independently combined with the signal detection false alarm rate (determined by the old/new criterion) to arrive at the total false alarm rate for each subject. Tables 1

and 2 show the results when the false remember rate was set at 15% and 30% of the overall remember rate. The primary effect of our increasing the false remember rate was the reduction in the suppression effect. Thus, as the false alarm rate became increasingly determined by the same two processes that affect the hit rate, the role of the correct remember rate in the regression naturally became more redundant. At a false remember rate of 15%, the suppression effect was still clear, and the pattern of regression estimates and correlations still resembled the actual aggregate data, in which the false remember rate was 13% of the overall remember rate. By the time the false remember rate reached 30%, the relationship between the hit rate and correct remember rate was no longer sufficiently unique (in comparison with the relation between correct remembers and false alarms) to demonstrate suppression, and redundancy became evident in the regression and correlation data. This outcome suggests that one should expect the degree or level of suppression observed in actual data to be inversely related to the level of false remember reports.

The role of guessing. Recently, Gardiner, Ramponi, and Richardson-Klavehn (1998) presented recognition data in which subjects were allowed to make three types of item endorsements: *remember*, *know*, and *guessing*. The results indicated that when the subjects reported that they had guessed, the odds of success were approximately even, in contrast to remember and know responses, which were both considerably higher. This outcome was taken by Gardiner et al. as suggesting that guess responses may be qualitatively different from both remember and know responses.

In the present study, we have assumed that instances of guessing are already appropriately modeled within the signal detection framework, under the standard assumption that as a given item's familiarity approaches the criterion for responding *old*, subjects will become increasingly uncertain of their endorsements (for a similar argument, see Hirshman, 1998). From a signal detection perspective, it is not surprising that the odds of success decline and approach even for the lowest levels of confidence, provided that the subject is largely unbiased (i.e., old/new criterion is centered midway between the distributions). In the case of a group of subjects who are on average unbiased (viz., some liberal, some neutral, and some conservative), the experiment-wide odds of successful guessing responses will also approach even. To illustrate, if we take the conservative and liberal observers illustrated in Figure 1, and assume that they will respond *guessing old* whenever items fall within .25 standard deviations above their old/new criterion, we can use the relative areas under the distributions to calculate their odds of success. For the conservative observer, whose criterion is set at 1.5, the odds are 3.7 to 1. For the liberal observer, whose criterion is set at $-.25$, the odds are .27 to 1 that he or she will be correct, or 3.7 to 1 that he or she will be mistaken (i.e., below chance). Thus, despite the fact that each subject has the same memory capability, their cutoff values for re-

sponding “old” greatly affects their raw odds of success. Furthermore, if the data were averaged across the 2 observers, the experiment-wide odds of successful guessing responses would then be even, despite the fact that the individual odds of successful guessing are reciprocal. These factors illustrate the difficulty in interpreting raw odds data as being indicative of different mnemonic response types. Modeling guessing on the same confidence continuum as other non-recollective endorsements arguably avoids such problems and is consistent with the observation that when guessing, the odds of success can be quite low.

DISCUSSION

Overall, the results pose a significant challenge to the two-criterion signal detection model. Although this model is capable of reproducing observed correlations between remember and hit rates, and between hit and false alarm rates, it makes errant predictions in regard to the relation between remembering and committing false alarms. Under the model there should be an observed (although indirect) correlation between remembering and false alarm rates in the raw scores. Observers with lax old/new criteria should have elevated false alarm rates and a greater tendency to report remembering. The reverse should be true for conservative observers. This positive relationship is clear and statistically reliable in the generated data (see the right panel of Figure 2B). In contrast, the experimental data suggest no such relationship between remember and false alarm rates (see the right panel of Figure 2A). The second prediction of the two-criterion signal detection model is that in the presence of hit rate information, the remember rate will not contribute to predicting false alarms (i.e., redundancy). Because multiple regression extracts the unique contribution of each of the predictors, and remember rates have no unique information (from the two-criterion signal detection viewpoint), a null contribution is predicted. This is indeed what occurred in the Monte Carlo simulation; however, this prediction stands in contrast to the actual experiment results, which showed a large suppression effect that demonstrated the unique relationships between remembering and hits, and between hits and false alarms. These findings contradict the notion that “remember responses represent nothing more than conservative yes responses” (Donaldson, 1996, p. 524).

Although these findings pose problems for the two-criterion signal detection model, they are easy to interpret within a dual-process signal detection framework (Yonelinas, 1994). In the present group of experiments, remember rates had a largely unique relation with hit rates because new items rarely triggered the retrieval of contextual information about the previous study event (for exceptions, see Dobbins, Kroll, & Liu, 1998; Roediger & McDermott, 1995); hence, the false alarm rate was relatively free of the recollection process. In contrast, a subject’s familiarity response criterion determines both

nonrecollective hits and false alarms, and when the effects of recollection are partialled or removed from the hit rate through regression, this relationship is clarified (i.e., the beta weight is higher than the raw correlation). This conclusion is further strengthened by the Monte Carlo simulation of the dual-process model, which reproduced both the patterns of correlations and the suppression effect that was observed in the experimental data.

These findings highlight the utility of examining individual subject data with regression techniques when contrasting recognition models.⁶ The failure of the signal detection model to predict basic patterns of responding across individual subjects brings into question its utility in describing group-level differences and shows that it is in need of revision. The dual-process signal detection model does this by incorporating an additional contextual retrieval process. We do not deny that studied items often engender only a feeling of familiarity during testing, and that such information is well modeled as being continuous in nature, using signal detection theory. Although such feelings of familiarity can be used by the subject to infer that an item was in fact studied, this is clearly a more distant inference than that which would be required had the subject remembered something uniquely related to the study context—for example, the subject’s having mispronounced the item when he/she had previously read it on the computer monitor. Given this premise, the attempt to model remember-know performance entirely within the signal detection framework amounts to the assertion that it is appropriate to scale responses that are accompanied by the retrieval of contextual information identically to those that are not. Such an equivalency of scale challenges the notion that subjects are capable of remembering distinct personal episodes when appropriately cued and is inconsistent with the observed relations among phenomenological reports of individual subject data.

REFERENCES

- CONGER, A. J. (1974). A revised definition for suppressor variables: A guide to their identification and interpretation. *Educational & Psychological Measurement*, *34*, 35-46.
- DOBBINS, I. G. (2000). *The systematic discrepancy between A' for overall recognition and remembering: A dual process account*. Manuscript submitted for publication.
- DOBBINS, I. G., KROLL, N. E. A., & LIU, Q. (1998). Confidence-Accuracy inversions in scene recognition: A remember-know analysis. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, *24*, 1306-1315.
- 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 Sciences*, *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.
- GARDINER, J. M., RAMPONI, C., & RICHARDSON-KLAVEHN, A. (1998).

- Experiences of remembering, knowing, and guessing. *Consciousness & Cognition*, *7*, 1-26.
- HIRSHMAN, E. (1998). On the utility of the signal detection model of the remember-know paradigm. *Consciousness & Cognition*, *7*, 103-107.
- 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, FL: Harcourt Brace.
- RAJARAM, S., & ROEDIGER, H. L., III. (1996). Remembering and knowing as states of consciousness during retrieval. In J. Cohen & J. W. Schooler (Eds.), *Scientific approaches to consciousness* (pp. 213-240). Mahwah, NJ: Erlbaum.
- RATCLIFF, R., SHEU, C.-F., & GRONLUND, S. D. (1992). Testing global memory models using ROC curves. *Psychological Review*, *99*, 518-535.
- ROEDIGER, H. L., III, & McDERMOTT, K. B. (1995). Creating false memories: Remembering words not presented in lists. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, *21*, 803-814.
- ROEDIGER, H. L., III, McDERMOTT, K. B., & ROBINSON, K. (1998). The role of associative processes in creating false memories. In M. A. Conway, S. E. Gathercole, & C. Cornoldi (Eds.), *Theories of Memory: Volume II* (pp. 187-245). East Sussex: Psychology Press.
- STRETCH, V., & WIXTED, J. T. (1998). On the difference between strength-based and frequency-based mirror effects in recognition memory. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, *24*, 1379-1396.
- 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.
- WIGGINS, J. S. (1973). *Personality and prediction: Principles of personality assessment*. Reading, MA: Addison Wesley.
- 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., 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. Alternately, one might suppose that subjects adopt an optimal old/new response criterion midway between the old- and new-item distributions. This would mean that subjects with a higher d' would have lower false alarm rates. In addition, because the optimal criterion location "lags" behind the mode of the old-item distribution as d' increases, these subjects would also have higher hits rates. Thus, what is predicted is a mirror effect in the individual subject data (see Stretch & Wixted, 1998, for further discussion). However, there was no evidence for such a placement strategy in the present data.

2. Throughout the paper we use the terms *remember rates* and *correct remember rates*, interchangeably. Incorrect remember responses are explicitly labeled as such.

3. Donaldson (1996) used both d' and A' when he compared the accuracy of remember responses with that of overall recognition on the basis of the claim that remember responses arise from the same distributions that govern overall responding. In order to simulate this model, one must specify an appropriate familiarity distribution from which to sample; we chose to use the normal distribution with its relative distance measure d' . However, in the case of A' , Macmillan and Creelman (1996) demonstrated that the implied relative operating characteristic (ROC) of the A' metric is not consistent with a known single type of underlying familiarity distribution, but instead "morphs" from one that is consistent with logistic distributions at low performance levels to one that is consistent with double equal threshold, rectangular distributions at high performance levels. Given the similarity in shape between logistic and normal distributions, the present data and conclusions would be largely unaffected by the use of A' at low performance levels. However, the fact that at high performance levels the implied ROC for A' resembles a double equal threshold process is clearly at odds with the form of empirical ROCs for single item, old/new recognition (see Macmillan & Creelman, 1991; Swets, 1986). This precluded its use in the present simulations.

It is also worth noting that in the present signal detection simulation data, where d' remember and d' recognition were identical by construction, A' remember was systematically lower than A' recognition (57 out of 72 cases, $p < .001$, sign test). Such a pattern was evident in the meta-analysis data of Donaldson (1996) in 60 out of 80 comparisons and has also been observed in empirical individual subject data (e.g., Gardiner & Gregg, 1997). Overall, these findings suggest that the A' metric may be systematically biased (Dobbins, 2000).

4. The actual d' in the experimental data was 0.93 when calculated from the overall hit and false alarm rates and 1.05 ($\sigma = .55$) when the rates were averaged across subjects.

5. Additional simulations were conducted in which the remember criterion was sampled from a uniform distribution between zero and two standard deviations above the old/new criterion. The pattern of results was unchanged.

6. We thank John Gardiner for pointing out that such regression techniques may also be useful in traditional meta-analysis studies, provided that enough data points are available for reliable regression, in which case the hit, remember, and false alarm rates of each experiment would serve as the units of analysis.

(Manuscript received February 17, 1999;
revision accepted for publication March 10, 2000.)