

Adaptive changes of response criterion in recognition memory

EVAN HEIT, NOELLIE BROCKDORFF, and KOEN LAMBERTS
University of Warwick, Coventry, England

An experiment was conducted to investigate people's ability to vary a response criterion strategically in a recognition memory task, as a function of the length of time given to process the test stimuli (from 100 to 1,500 msec). The experiment used the response signal procedure, in which the participants responded after a signal that came at a variable time delay from stimulus onset. The proportion of new versus old test items was varied systematically with the time of the response signal, with the proportion of new test items rising, falling, or staying constant at later signals. It was found that the participants' response biases changed adaptively, becoming more conservative at later signals in the rising condition, becoming less conservative in the falling condition, and not changing significantly in the constant condition. Theoretical and methodological implications for recognition memory research are discussed.

Memory researchers have increasingly become interested in the time course of recognition memory—that is, how recognition judgments are created over time. Collecting data over the course of the creation of a recognition judgment should provide a detailed picture of recognition that is especially informative for developing models of recognition memory, going beyond data that simply reflect the end point of the judgment process. Relatedly, time course data should be especially useful for addressing such questions as whether there are multiple processes involved in recognition memory (e.g., Brockdorff & Lamberts, 2000; Rotello & Heit, 1999) and whether different sources of information are available at different times (e.g., Doshier & Rosedale, 1991).

One technique for looking at the time course of recognition is to measure event-related potentials (ERPs), electromagnetic signals in the brain at different intervals after a stimulus presentation (e.g., Curran, 2000). However, it is also possible to look at the time course of recognition with behavioral measures alone, using the response signal technique (introduced by Reed, 1973). With this technique, the participant is interrupted at various points in time while making a recognition judgment. The participant is instructed to respond quickly after the interrupting signal (e.g., a tone), making a judgment on the basis of whatever assessment has been completed up to that point. The response signal comes at some point from, say, 100 to 1,500 msec from stimulus onset, determined randomly for each judgment trial.

With the response signal technique, a number of interesting results in recognition memory have been obtained.

For example, Doshier and Rosedale (1991), looking at memory for word pairs, found that retrieving episodic information (whether the words had been studied together) had a different time course than did retrieving semantic information (whether the words were related). This technique has also been applied to comparisons of the time course of recognition versus lexical decision (Hintzman & Curran, 1997), recognition versus modality judgment (Hintzman & Caulton, 1997; McElree, Dolan, & Jacoby, 1999), item recognition versus associative recognition (Gronlund & Ratcliff, 1989; Ratcliff & McKoon, 1989), and item recognition versus list discrimination (Hintzman, Caulton, & Levitin, 1998; Rotello & Heit, 2000).

Taken together, experiments with the response signal technique have used a variety of statistical measures of recognition memory performance. Virtually all the experiments have reported a combination of raw scores (hit rates and false alarms) and derived scores (such as d'). In general, we support the idea of using multiple measures of performance, targeting different aspects of the data and giving a full picture of the results.

However, different measures of performance may give conflicting results. In experiments on word recognition memory, Hintzman and Curran (1994) reported a striking nonmonotonic pattern of results for foil test stimuli that were similar to old training items—for example, for FROGS when FROG had been studied. The false alarm rate on these foil items rose up to around 700 msec, then fell with additional processing time. This pattern was interpreted in terms of a second, slower process having an effect on recognition later in processing, helping people to reject similar foils. However, Rotello and Heit (1999) reanalyzed these data, using derived scores (specifically dL , the logistic version of d'), and found that the pattern of performance for similar foils was a monotonic rise to an asymptote, rather than being “peaked.” This dissociation between raw scores and derived measures can also go the opposite way. In an

This research was supported by a grant from the Biotechnology and Biological Sciences Research Council of the United Kingdom. Part of this research was presented at the 2000 Annual Meeting of the Psychonomic Society, New Orleans. Please address correspondence to E. Heit, Department of Psychology, University of Warwick, Coventry CV4 7AL, England (e-mail: e.heit@warwick.ac.uk).

experiment on list-specific recognition memory for words, Rotello and Heit (2000, Experiment 3) found that raw scores reached an asymptote, then stayed flat, but that derived scores showed a peaked pattern. Therefore, it does seem important to consider multiple measures of performance, in case different measures might suggest different conclusions.

In addition, there is potentially a problem in using raw scores, particularly with the response signal technique. It is widely accepted that raw scores, such as hit rate and false alarm rate, reflect a mixture of response to the stimulus itself, as well as guessing or other response biases (e.g., Macmillan & Creelman, 1991). The conventional way to overcome the problem of raw scores is to use derived measures, such as d' , that are meant to correct for possibly varying response bias. If participants in a response signal experiment have different response criteria at different points in time, looking at raw scores could be rather misleading. For example, if participants have an overall tendency to respond *new* at early test signals and to respond *old* at later test signals, it would be difficult to compare, say, a false alarm rate of 60% at 100 msec with a false alarm rate of 40% at 1,500 msec.

Before describing some results that address whether response criteria vary in response signal experiments, we present two intuitive arguments. On the one side, one might argue that because the response signal task moves so quickly, interrupting participants at random, unpredictable times, it is implausible that participants could strategically vary their response criteria. The argument is that participants cannot choose to have different criteria for different stimuli; instead, response criterion is an overall characteristic of performance in a recognition experiment. On the other side, the argument would be that responding at early and at late signals are, in effect, very different tasks. Making a recognition judgment after 100 msec is so different from making a recognition judgment after 1,500 msec that it is plausible that different criteria would be in effect. Another way to put this argument is that with five to eight different response signals between 100 and 1,500 msec, it would be difficult for participants to maintain the same response criteria for different time signals. Next, we will turn to past results.

Hintzman, Caulton, and Curran (1994) reported three word recognition memory experiments that used the response signal method. For all three experiments, analyses of variance (ANOVAs) did not show significant changes in response bias (using cL , the logistic equivalent of the c measure of bias) as a function of time. The graphs of the data showed a rather flat average cL , not varying for different response signals. At first inspection, there was not much evidence for a changing response bias. However, Hintzman et al. (1994) noted the high variability of their cL measures. This variability could reduce the power to detect a small change in criterion. Also, the pattern of data was consistent with individual participants varying their biases as a function of time, but with different participants adopting different strategies over time. (Indeed, Hintzman

et al., 1994, reported that individual participants' response biases were correlated from one response signal to another, fitting the idea that there were systematic individual differences.) In sum, Hintzman et al.'s (1994) results do not point directly to changes in criteria, but as null results, they do not rule this out, either.

Rotello and Heit (1999) briefly described (footnote 8) several analyses of response bias; they reported some significant changes in response bias for their own experiments and those of Hintzman and Curran (1994). The general finding in Rotello and Heit's (1999) experiments was a more conservative response bias at later time signals, when more information would be available to the participant. Therefore, Rotello and Heit (1999) argued that for response signal experiments, raw scores need to be interpreted with caution, because changes in false alarm rate could reflect mere changes in response bias. However, they did not provide much detail about these analyses. Also, Rotello and Heit (1999) did not make the crucial distinction between response bias and response criterion. According to signal detection theory (e.g., Macmillan & Creelman, 1991), the c measure of response bias denotes the position of the decision criterion relative to the halfway point between the new and the old distributions. As has been highlighted in recent articles by Miller and Wolford (1999), Wickens and Hirshman (2000), and Wixted and Stretch (2000), it is problematic to treat c as an absolute measure of response criterion when comparing results from two experimental conditions. (There are analogous problems for the cL measure of bias.) If the two experimental conditions also show different levels of sensitivity (d'), the midpoint between distributions will be in a different place. In response signal experiments, performance in terms of d' typically varies at different response signals. Hence, it would not be straightforward to compare measures of bias at different times, and even if a change in bias is manifested, it might not reflect a true change in underlying response criterion.

Rationale of Experiment

In our experiment, we created more suitable conditions for investigating whether people can vary the response criterion strategically in a response signal task. The key idea was to manipulate the pattern of test stimuli to encourage different patterns of response bias. Several previous studies (e.g., [Estes & Maddox, 1995](#); [Hirshman & Henzler, 1998](#); [Ratcliff, Sheu, & Gronlund, 1992](#)) have suggested that, more generally in recognition experiments, response bias is sensitive to the proportion of new versus old test items in an adaptive manner. For example, people seem to show a more conservative response bias when there is a high proportion of true *new* items on a test list. However, these experiments did not use the response signal technique; they focused on comparisons between groups of participants or between lists, in which case a participant was expected to adopt the same bias for all the items on a test list. In our experiment, we tested different proportions of new and old items at different time signals, so that the

participants might adopt different response biases at different time signals if they were able to do so. The configuration of test items was varied between groups of participants; therefore, we expected to find systematic differences in patterns of response bias between groups. It was also expected that different groups of participants would be roughly equivalent in term of sensitivity measures, so that comparing the response bias measures between groups would be fairly straightforward, making it easier to draw inferences about true differences in criterion.

We compared two main conditions in which the proportion of new items tested varied as a function of processing time. In the *rising* condition, the proportion of new items tested was low at early response signals, but this proportion increased at later signals. In the *falling* condition, the proportion of new items tested was high at early response signals, falling at later response signals. We predicted that the participants in the two groups would show different patterns of response bias over processing time, reflecting different strategies based on the distribution of new and old test items at different signals. For comparison, we also included a third condition in which the proportion of new test items was *constant* across response signals.

Method

Participants. The paid participants were 18 students from the University of Warwick.

Apparatus and Stimuli. The experiment was controlled by a Pentium 400-MHz computer with a 17-in. color monitor. Responses were registered by means of two microswitches connected to the computer's parallel port.

The picture stimuli consisted of color photographs, presented in the middle of the screen on a black background. (For other recent experiments in which the response signal technique was used to study picture recognition memory, see Brockdorff & Lamberts, 2000, and Lamberts, Brockdorff, & Heit, 2002.) The 360 photographs measured 9.2×9.2 cm, showed a wide variety of outdoor scenes, and looked much like postcards. For each participant, the 360 stimuli were randomly assigned to six study/test blocks. Each study list consisted of 30 experimental pictures and 6 filler pictures, 3 of which were used at the start of the list and 3 at the end of the list, to minimize primacy and recency effects. Each test list consisted of 60 experimental pictures, 30 old studied items and 30 new unstudied items.

Procedure. Each participant attended one session, lasting about 70 min. Each session consisted of a practice stage and six old–new recognition memory tests in which a response signal procedure was used. Each memory test consisted of a study phase followed immediately by a test phase.

In each study phase, all 36 pictures were presented on the computer screen, 1 picture at a time, at a rate of 3 sec per picture. Immediately following the study phase, a recognition test was given on the 30 old and 30 new pictures. On each test trial, a cue (a cross) was shown at the center of the screen for 500 msec. The screen went blank (black) for 100 msec, and then a picture appeared. At variable time lags after the stimulus appeared on screen (100, 250, 400, 600, 900, or 1,500 msec), a 1000-Hz tone sounded, and the stimulus was replaced on screen by a mask made up of small multicolored squares. These time lags cover the range that has typically been used in other response signal studies of recognition. The participants were instructed to make a recognition judgment immediately after they heard the tone and to respond as accurately as possible. The participants were instructed to press the *old* key if the test item was one

previously seen during study and to press the *new* key if the item had not been seen. The response button mappings were counterbalanced across participants. If no response was made within 300 msec of the onset of the tone or if a response was made before the onset of the tone, an appropriate error message was displayed. In addition, summary feedback about accuracy and proportion within time was given at the end of each test block.

The participants were assigned to three different conditions, with 6 participants in each condition. The study phases in the three conditions were identical, but the three conditions differed in the test phase. Each test block in all three conditions consisted of 60 trials. In the rising condition, the ratio of new stimuli to old stimuli was manipulated across time lags. In the two earliest time lags, the ratio of new stimuli to old stimuli was one new to nine old; in the two middle time lags, the ratio was five new to five old; and in the last two time lags, the ratio was nine new to one old. In the falling condition, for the two earliest time lags, the ratio of new stimuli to old stimuli was 9 new to one old; in the two middle time lags, the ratio was five new to five old; and in the last two time lags, the ratio was one new to nine old. Finally, in the constant condition, equal numbers of pictures of each type were assigned to each response signal condition (five new and five old in each time lag; see Table 1 for a summary). The participants were not informed that the ratios of new versus old test items might vary depending on the time before the response signal.

Before the experimental blocks, the participants were given a practice block, which was exactly like an experimental study/test block except that the stimuli were color drawings, rather than photographs. In the practice test phase, equal numbers of pictures of each type (new or old) were assigned to each response signal condition.

Results

The data were trimmed to remove responses in which the participants either failed to respond within 350 msec of a response signal (10%) or made anticipatory responses earlier than 100 msec after the response signal (2%).

The mean accuracy over all the participants on each condition is shown in Figure 1. (Note that the *x*-axis reflects time to signal plus average response time.) Accuracy is expressed in terms of the d' index of sensitivity. To avoid undefined values of d' , in the computation of d' , the hit rates and the false alarm rates were adjusted by adding .5 to the number of *old* responses and dividing by the number of responses +1 (Snodgrass & Corwin, 1988). The clear results are that d' increased over the course of judgment, reaching a high level of performance, and that d' did not vary as a function of condition (rising, falling, or constant). The conclusions were supported by an ANOVA. There was a significant main effect of time of response signal [$F(5,75) = 82.42, MS_e = 0.30, p < .001$]. The effect of condition was not significant [$F(2,15) = 0.96, MS_e = 1.41$], and likewise, the interaction between con-

Table 1
Ratio of New to Old Test Stimuli Across Time Lags
in the Three Test Conditions

Condition	Time Lag (msec)					
	100	250	400	600	900	1,500
Rising	1:9	1:9	5:5	5:5	9:1	9:1
Falling	9:1	9:1	5:5	5:5	1:9	1:9
Constant	5:5	5:5	5:5	5:5	5:5	5:5

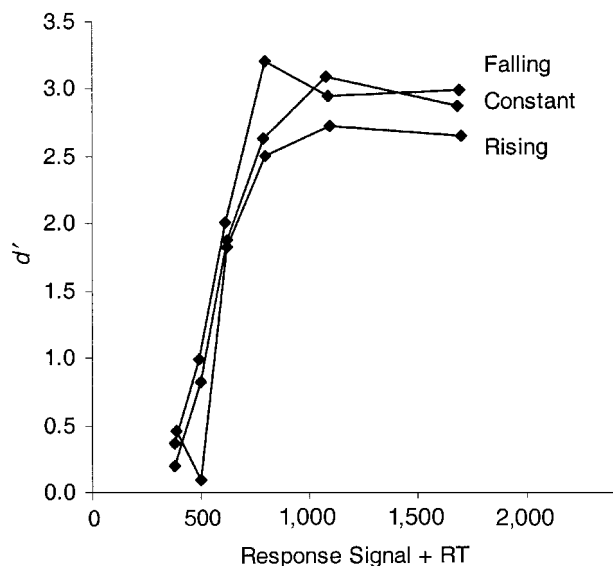


Figure 1. Sensitivity (d') at varying response signals.

dition and response signal was not significant [$F(10,75) = 1.06$, $MS_e = 0.30$].

For clarity, the bias results are presented separately for the constant condition and for the two other (rising and falling) conditions. The top panel of Figure 2 shows response bias for the constant condition in terms of the c measure of bias. Note that higher values of c indicate more conservatism—that is, a greater tendency to say *new* (Macmillan & Creelman, 1991, p. 33). (In the figure, error bars show standard errors.) In an ANOVA for just the constant condition, the effect of response signal time did not reach the level of statistical significance [$F(5,25) = 0.36$, $MS_e = 0.15$]. Next, the bottom panel of Figure 2 shows response bias for the rising and falling conditions. In the rising condition, the participants became more conservative at later response signals as the proportion of new test items increased. Likewise, in the falling condition, the participants became less conservative at later response signals as the proportion of new test items decreased. Critically, an ANOVA for these two conditions showed a significant interaction between condition and response signal time [$F(5,50) = 11.46$, $MS_e = 0.10$, $p < .001$]. There was no significant main effect of condition [$F(1,10) = 0.00$, $MS_e = 0.67$] or response signal [$F(5,50) = 1.33$, $MS_e < 0.10$], however.

Finally, for completeness, in Table 2 we report the raw hit rates and false alarm rates. Perhaps most notable is that the false alarm rates generally decrease at later response signals. This finding was originally taken by Rotello and Heit (1999) to suggest changes in response bias over the time course of making a recognition judgment.

Discussion

The results very clearly showed that people can adjust their response biases strategically at different time lags when the response signal procedure is used. In the rising

condition, in which it would be adaptive to increasingly give *new* responses at longer lags, to reflect the greater proportion of new test items, people did become more conservative over time. Likewise, the participants adaptively changed their response bias in the falling condition, reflecting the decreasing proportion of new items over time.

Drawing inferences about response criterion. We next turn to the question of whether these clear changes in the c measure of response bias can be interpreted as truly reflecting shifts in the underlying criterion. We note that there was no significant main effect of condition on d' and, likewise, the interaction between condition and response signal did not have a significant effect on d' . Hence, the values of d' at a particular response signal should be comparable from one condition to another. In terms of signal detection theory, the distributions of old items should be the same in all three conditions. Thus, at any particular response time, the response bias differences can be safely interpreted as differences in criterion from one condition to another. Therefore, we would conclude that the different patterns of testing for the rising and the falling conditions led to different patterns of response criterion.

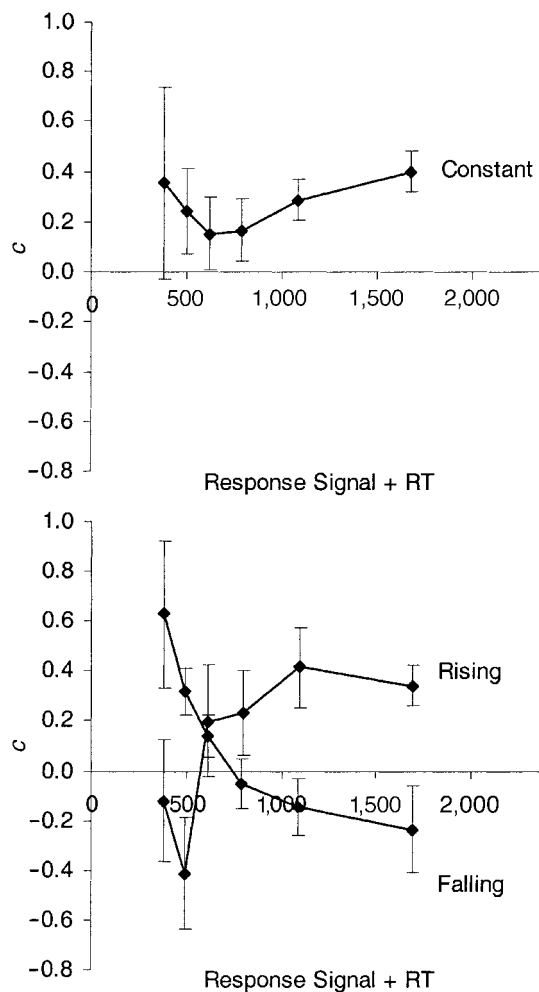


Figure 2. Response bias (c) at varying response signals.

Table 2
Hit Rate and False Alarm Rate Across Time Lags
in the Three Test Conditions

Condition	Time Lag (msec)					
	100	250	400	600	900	1,500
	Hit Rate					
Rising	.66	.68	.76	.83	.85	.88
Falling	.28	.58	.80	.95	.93	.95
Constant	.30	.56	.77	.87	.88	.83
	False Alarm Rate					
Rising	.46	.64	.16	.07	.05	.07
Falling	.21	.21	.14	.06	.03	.06
Constant	.31	.26	.14	.06	.02	.04

A perhaps more conservative analysis would reflect that although there were no significant effects of condition on d' , the graph in Figure 1 does suggest some slight, albeit nonsignificant, variations in d' from one condition to another. If these small differences were real, they could make the c measures of bias slightly misaligned, because the midpoint between the new and the old distributions could differ slightly between conditions. In an attempt to address this possibility, we used each participant's d' measure to rescale his or her c measure, so that c was measured relative to a zero point, rather than relative to a midpoint between distributions. Using these rescaled c measures, we were able to draw the same main conclusions. That is, at early time signals, the participants in the falling condition were more conservative than the participants in the rising condition, and at later time signals, the participants in the rising condition were more conservative than the participants in the falling condition.

Having established that response criterion truly differed between conditions, we next turn to the question of whether the participants truly varied their response criteria within a condition. This is a more difficult question, because there were substantial changes in d' over the time course of making a recognition judgment. Therefore, the c measures of bias must be interpreted with caution when one is looking at different time lags within a condition. However, we note that c rose over time in the rising condition and fell over time in the falling condition. Any shifts in the distributions over time should affect the rising and falling conditions in the same direction, so they could not explain the different changes in the two conditions. Hence, there must be some true shifts in the response criterion over time in this experiment, taking the results from all the conditions together.

Conclusions. We see these results as having methodological as well as theoretical implications. Practically, as was argued by Rotello and Heit (1999), when there is potential for participants to vary their response criteria strategically, in response signal experiments or when test lists differ in composition, raw score measures that do not correct for response bias should be interpreted with caution. Admittedly, most memory research already does supplement raw score measures with signal detection analyses,

but these results provide an even stronger motivation, particularly for response signal experiments. In cases in which raw scores and derived scores lead to different theoretical conclusions (e.g., Rotello & Heit, 1999, 2000), it should be noted that raw scores could well be influenced by changes in response criteria.

In theoretical terms, we note that the general debate about whether and when people can vary response criteria adaptively has implications for a number of theoretical issues in recognition memory (e.g., Estes & Maddox, 1995; Gallo, Roediger, & McDermott, 2001; Gillund & Shiffrin, 1984; Miller & Wolford, 1999; Rotello & Heit, 1999, 2000; Stretch & Wixted, 1998). We see our own results as documenting this ability, whereas past conclusions have been mixed. Indeed, we see striking parallels between our participants' ability to learn about the recognition test set and the phenomenon of unsupervised category learning, in which people learn about the structure of a stimulus set without detailed trial-and-error feedback (for a review, see Love, 2002). To give one dramatic example of unsupervised category learning, Palmeri and Flanery (1999) reported good performance on a multidimensional perceptual categorization test by participants who had not observed any training items at all, because the structure of the stimulus set could be inferred from the test list.

In general terms, there is a history of bringing together models of recognition memory and categorization (e.g., Estes, 1994). In recent work (Brockdorff & Lamberts, 2000; Lamberts et al., 2002), we have applied an exemplar model, originally designed for time course phenomena in categorization, to experiments on the time course of recognition memory. This model postulates a feature-sampling process that takes place over time, so that later judgments reflect more featural information. Generally speaking, this model can predict changes in response bias over the time course of making a recognition judgment. For example, the model would account for the general trend for decreased false alarm rates over time (see Table 2) by assuming that with more feature information available, new test items should be increasingly dissimilar to old, studied items. Moreover, this model has a means for modulating response bias, by varying the relative influences of old, studied items versus "background" memory traces from outside of the study list. What the model still lacks is a mechanism for learning appropriate response biases from the test list (but see Estes & Maddox, 1995, p. 1089, for relevant suggestions).

To conclude, we believe that greater attention to the issue of changes in response criterion should have a beneficial impact on future theoretical development in recognition memory research.

REFERENCES

- BROCKDORFF, N., & LAMBERTS, K. (2000). A feature-sampling account of the time course of old-new recognition judgments. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, *26*, 77-102.
- CURRAN, T. (2000). Brain potentials of recollection and familiarity. *Memory & Cognition*, *28*, 923-938.

- DOSHER, B. A., & ROSEDALE, G. (1991). Judgments of semantic and episodic relatedness: Common time-course and failure of segregation. *Journal of Memory & Language*, **30**, 125-160.
- ESTES, W. K. (1994). *Classification and cognition*. New York: Oxford University Press.
- ESTES, W. K., & MADDOX, W. T. (1995). Interactions of stimulus attributes, base rates, and feedback in recognition. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **21**, 1075-1095.
- GALLO, D. A., ROEDIGER, H. L., III, & McDERMOTT, K. B. (2001). Associative false recognition occurs without strategic criterion shifts. *Psychonomic Bulletin & Review*, **8**, 579-586.
- GILLUND, G., & SHIFFRIN, R. M. (1984). A retrieval model for both recognition and recall. *Psychological Review*, **91**, 1-67.
- GRONLUND, S. D., & RATCLIFF, R. (1989). Time course of item and associative information: Implications for global memory models. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **15**, 846-858.
- HINTZMAN, D. L., & CAULTON, D. A. (1997). Recognition memory and modality judgments: A comparison of retrieval dynamics. *Journal of Memory & Language*, **37**, 1-23.
- HINTZMAN, D. L., CAULTON, D. A., & CURRAN, T. (1994). Retrieval constraints and the mirror effect. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **20**, 275-289.
- HINTZMAN, D. L., CAULTON, D. A., & LEVITIN, D. J. (1998). Retrieval dynamics in recognition and list discrimination: Further evidence of separate processes of familiarity and recall. *Memory & Cognition*, **26**, 449-462.
- HINTZMAN, D. L., & CURRAN, T. (1994). Retrieval dynamics of recognition and frequency judgments: Evidence for separate processes of familiarity and recall. *Journal of Memory & Language*, **33**, 1-18.
- HINTZMAN, D. L., & CURRAN, T. (1997). Comparing retrieval dynamics in recognition memory and lexical decision. *Journal of Experimental Psychology: General*, **126**, 228-247.
- HIRSHMAN, E., & HENZLER, A. (1998). The role of decision processes in conscious recollection. *Psychological Science*, **8**, 61-65.
- LAMBERTS, K., BROCKDORFF, N., & HEIT, E. (2002). Perceptual processes in matching and recognition of complex pictures. *Journal of Experimental Psychology: Human Perception & Performance*, **28**, 1176-1191.
- LOVE, B. C. (2002). Comparing supervised and unsupervised category learning. *Psychonomic Bulletin & Review*, **9**, 829-835.
- MACMILLAN, N. A., & CREELMAN, C. D. (1991). *Detection theory: A user's guide*. Cambridge: Cambridge University Press.
- MCELREE, B., DOLAN, P. O., & JACOBY, L. L. (1999). Isolating the contributions of familiarity and source information to item recognition: A time-course analysis. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **25**, 563-582.
- MILLER, M. B., & WOLFORD, G. L. (1999). The role of criterion shift in false memory. *Psychological Review*, **106**, 398-405.
- PALMERI, T. J., & FLANERY, M. A. (1999). Learning about categories in the absence of training: Profound amnesia and the relationship between perceptual categorization and recognition memory. *Psychological Science*, **10**, 526-530.
- RATCLIFF, R., & MCKOON, G. (1989). Similarity information versus relational information: Differences in the time course of retrieval. *Cognitive Psychology*, **21**, 139-155.
- RATCLIFF, R., SHEU, C.-F., & GRONLUND, S. D. (1992). Testing global memory models using ROC curves. *Psychological Review*, **99**, 518-535.
- REED, A. V. (1973). Speed-accuracy trade-off in recognition memory. *Science*, **181**, 574-576.
- ROTELLO, C. M., & HEIT, E. (1999). Two-process models of recognition memory: Evidence for recall-to-reject? *Journal of Memory & Language*, **40**, 432-453.
- ROTELLO, C. M., & HEIT, E. (2000). Associative recognition: A case of recall-to-reject processing. *Memory & Cognition*, **28**, 907-922.
- SNODGRASS, J. G., & CORWIN, J. (1988). Pragmatics of measuring recognition memory: Applications to dementia and amnesia. *Journal of Experimental Psychology: General*, **117**, 34-50.
- 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**, 137-196.
- WICKENS, T. D., & HIRSHMAN, E. (2000). False memories and signal detection theory: Comment on Miller and Wolford (1999) and Roediger and McDermott (1999). *Psychological Review*, **107**, 377-383.
- WIXTED, J. T., & STRETCH, V. (2000). The case against a criterion-shift account of false memory. *Psychological Review*, **107**, 368-376.

(Manuscript received May 15, 2001;
revision accepted for publication May 31, 2002.)