The *remember* response: Subject to bias, graded, and not a process-pure indicator of recollection

CAREN M. ROTELLO, NEIL A. MACMILLAN, JOHN A. REEDER, and MUNGCHEN WONG
University of Massachusetts, Amherst, Massachusetts

Recognition memory judgments have long been assumed to depend on the contributions of two underlying processes: recollection and familiarity. We measured recollection with receiver-operating characteristic (ROC) data and remember–know judgments. Under standard remember–know instructions, the two estimates of recollection diverged. When subjects were told they might need to justify their *remember* responses to the experimenter, the two estimates were more likely to agree. The data support the conclusion that *remember* responses are generally based on a continuous underlying process but that specific task instructions can produce data that appear consistent with a high-threshold recollective process. Models based on signal detection theory provide a better account of these data than does the dual-process model (Yonelinas, 1994) or process-pure interpretations.

Recognition judgments are widely believed to be based on two underlying processes: recollection and familiarity. Here, we focus on two behavioral techniques for identifying these components: the remember–know paradigm and item recognition rating experiments.

In the remember–know paradigm (Tulving, 1985), subjects are asked to distinguish the subjective experience of *remembering* something specific about the occurrence of a study item from that of *knowing* that it was studied despite their failure to retrieve any specific details. A variety of empirical dissociations of remembering and knowing have been taken as evidence that the judgments are based on different underlying processes or memory systems (for a recent review, see Gardiner & Richardson-Klavehn, 2000). This *process-pure* interpretation underlies the majority of empirical remember–know studies.

An influential quantitative version of the process-pure hypothesis, the *dual-process model* (Yonelinas, 1994), proposes two converging operations for measuring the amount of recollection, which we denote R_{DP}. (1) In the remember–know paradigm, *remember* judgments draw only on recollection, *know* judgments only on familiarity (Yonelinas, 2001). R_{DP} should equal the proportion of old items that are remembered, but to allow for small amounts of *nonmemorial noise*, it is better estimated when the proportion of remembered new items is subtracted from that proportion (Yonelinas, 2002).[1] (2) In rating experiments,

receiver-operating characteristic (ROC) curves predicted by the model have a *y*-intercept that equals R_{DP}. Thus, ROC curves and *remember* judgments should produce essentially identical estimates of the underlying recollection process (Yonelinas, 2001, 2002).

Two studies from Yonelinas's laboratory that allowed estimation of R_{DP} from both *remember* judgments and ROC data simultaneously showed this equivalence (Yonelinas, 2001, Experiment 1; Yonelinas, Dobbins, Szymanski, Dhaliwal, & King, 1996, Experiment 2).[2] The subjects in these studies followed almost all highest confidence (and few lower confidence) *old* judgments with remember responses; know responses occurred at a variety of confidence levels. This pattern of responding leads directly to the similar estimates of R_{DP}.

In the only other experiments for which both ROCs and remember–know judgments were reported for the same subjects, a different pattern was found. Rotello, Macmillan, and Reeder (2004, Experiments 1A and 1B) observed remember judgments to words recognized with a variety of confidence levels. When the dual-process model is fit to these data, the estimates of R_{DP} from ROCs and from remember judgments diverge: In Experiment 1A, the ROC analysis concluded that R_{DP} equals .18, whereas the data from the remember judgments imply that R_{DP} is .26 (estimated from highest confidence responses) or .29 (overall); in Experiment 1B, the ROC implies that R_{DP} equals .09 and the data from the remember judgments indicate that R_{DP} is .23 (highest confidence) or .25 (overall). According to these data, if the ROC *y*-intercept measures recollection, the remember judgments do not. These results thus call into question the standard interpretation of remember–know data.

In Experiment 1, we attempted to replicate these contradictory results by varying the instructions provided to the
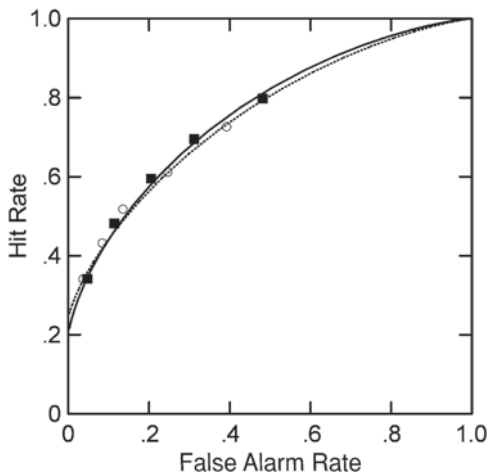
Figure 1. Item recognition ROCs for Experiment 1. Open circles represent the conservative condition, and closed squares represent the neutral condition. The superimposed functions were generated with the dual-process model (Yonelinas, 1994).

subjects. Rotello et al. (2004) followed standard instructions (Rajaram, 1993), in which any recollected aspect of the study experience merited a remember decision. In contrast, Yonelinas and his colleagues' instructions were more specific about what justified a remember judgment; subjects were told to reserve those responses for recollections they could explain to the experimenter. Our subjects studied a list of words and, at test, rated their *old–new* confidence and provided binary remember–know judgments under these two types of remember–know instructions.

EXPERIMENT 1

Method

Subjects. Fifty-two undergraduate students at the University of Massachusetts participated for extra credit in their psychology courses. All were native English speakers. Data were discarded for 2 subjects who failed to follow instructions, 1 who did not understand the remember–know distinction, and 1 who exhibited below-chance recognition performance. The remaining 48 subjects were randomly assigned to the conservative ($n = 24$) or neutral ($n = 24$) conditions.

Stimuli. One hundred twenty English nouns were selected from the MRC Psycholinguistic Database (Coltheart, 1981). The words were divided into two lists that were matched on number of syllables ($M = 1.5$), length ($M = 5.2$ letters), and frequency (Kučera & Francis, 1967; $M = 158$). Ten additional words were drawn from the same pool to serve as practice, primacy, and recency items.

Procedure. The experiment included a study phase, a practice phase during which the subjects' comprehension of the instructions was evaluated, and a test phase. In the study phase, 12 subjects in each condition studied one list of 60 words (randomly ordered), and 12 studied the other list. Four practice items, a primacy item, and a recency item were also included. Each word appeared in the center of a computer screen for 1,250 msec, with a 750-msec interstimulus interval. The subjects were instructed to study the words carefully in preparation for an upcoming memory test.

For each word, all the subjects rated their confidence that it had been on the study list, using a 6-point scale that was labeled at 3 points: (1) *sure new*, (2) *maybe studied*, to (6) *sure studied*. The subjects were asked to distribute their responses across all six ratings in order to indicate variations in confidence; they responded by pressing the appropriate number keys (1–6) on the computer keyboard. The second response, requested whenever the subjects reported any level of confidence that a word had been studied, was a binary judgment indicating whether they remembered or knew that they had studied the word. The instructions for this judgment were taken from standard definitions of remembering and knowing (Rajaram, 1993). In addition, an example was provided to distinguish between recognizing someone on the basis of remembered details of a past encounter and just knowing that someone was previously encountered on the basis of familiarity. In the conservative condition, the subjects were additionally instructed to respond *remember* only if they could actually describe specific details about the experience of studying a word (such as its visual appearance, its position in the list, or an image or association that it cued). The subjects responded *remember* or *know* by pressing the "z" or the "/" key, respectively. The subjects in the conservative condition also made a third judgment about words they remembered, indicating whether they recalled *one detail* (pressing "z") or *more than one detail* ("/") of the word's occurrence on the list. We did not analyze these final judgments.

Results

ROC curves for both groups are plotted in Figure 1. (Complete stimulus–response matrixes are given in Table A1 in the Appendix.) The curve for the neutral condition essentially overlaps that for the conservative condition but continues to higher values of *old* responses. When functions derived from dual-process theory are fit to the

Table 1
Estimates of R_{DP} for Both Experiments

Condition	List 1			List 2			Overall		
		Remember Rate			Remember Rate			Remember Rate	
	ROC Intercept	All	High Confidence	ROC Intercept	All	High Confidence	ROC Intercept	All	High Confidence
Experiment 1									
Neutral							.20	.32	.34
Conservative							.24	.27	.25
Experiment 2									
Voices/explicit–conservative	.33	.46	.42	.24	.32	.28	.31	.38	.35
Voices/explicit	.46	.57	.54	.25	.40	.37	.36	.49	.45
Voices/implicit	.32	.52	.45	.28	.33	.29	.31	.43	.37
No voices	.14	.55	.45	.17	.23	.19	.25	.39	.32

Note—All, remember hit rate − remember false alarm rate for all levels of confidence. High Confidence, remember hit rate − remember false alarm rate for highest level of confidence.

Table 2
Response Proportions (With Standard Deviations) From Experiment 1

Remember–Know Instructions	H		P(R|Old)		P(R|H)		F		P(R|New)		P(R|F)	
	M	SD	M	SD	M	SD	M	SD	M	SD	M	SD
Conservative	.73	.13	.33	.13	.45	.15	.39	.16	.05	.06	.14	.16
Neutral	.80	.12	.46	.17	.58	.18	.48	.17	.16	.12	.33	.25

Note—*H*, hit rate; *P(R\|old)*, *P*(remember\|old); *P(R\|H)*, *P*(remember\|hit); *F*, false alarm rate; *P(R\|new)*, *P*(remember\|new); *P(R\|F)*, *P*(remember\|false alarm). *H* and *F* are based on rating data with 1 = *new*; 2–5 = *old*.

curves (Yonelinas, 2002), the recollection parameter R_{DP} is estimated to be .24 in the conservative condition and .20 in the neutral condition (see Table 1). ROC estimates of R_{DP} should be similar across the two conditions, and they are.

The remember–know data are summarized in Table 2. The subjects in the conservative condition exhibited numerically lower *old* rates (hits, *H*, and false alarms, *F*), but the differences were not reliable. The effect of the instructions on remember (*R*) rates was evaluated by examining responses as a function of stimulus presentation (old or new) and recognition decision (*old* or *new*). Comparing the conservative condition with the neutral condition, remember responses to old and new items were significantly reduced [$P(R|old)$ = .33 vs. .46, $t(46)$ = 3.11, $p < .05$; $P(R|new)$ = .05 vs. .16, $t(46)$ = 3.78, $p < .01$], and the proportion of *old* decisions that were followed by a remember response was also reduced for both old [$P(R|H)$ = .45 vs. .58; $t(46)$ = 2.83, $p < .01$] and new [$P(R|F)$ = .14 vs. .33; $t(46)$ = 3.05, $p < .01$] items. Thus, our conservative instructions had the effect of reducing the subjects' willingness to respond *remember* once they recognized a word.

The parameter R_{DP} can be estimated either with all remember responses or just those made with highest confidence; both values are given in Table 1. In the key comparison between estimates obtained from ratings and from remember–know judgments, the two methods converge in the conservative condition (R_{DP} = .24 with the ROC approach and .25 or .27 using remember responses) but diverge considerably in the neutral condition (R_{DP} = .20 in the ROC and .34 or .32 using remember responses).

Some insight into this discrepancy is available in Figure 2, which shows that remember responding was greatly affected by instructions. In this figure, ROC curves are plotted for remember trials only: $P(R|old)$ is plotted against $P(R|new)$ as level of confidence varies, producing one curve per condition. The subjects in the conservative condition provided remember responses mostly to words recognized with highest confidence (*sure old*), and the remember ROC is very restricted. The neutral condition yielded quite different data, with remember responses spread over all levels of confidence. Under these instructions, the subjects reported remembering at graded levels of confidence, suggesting a continuum of recollection.

Discussion

The conservative condition data are well accounted for by the dual-process model. The subjects who were told

that they might need to explain their remember responses to the experimenter and who were asked to count their remembered details, did not produce many remember responses to words recognized with lower confidence. The proportion of remember false alarms, which should be close to zero according to the model, was .05. The results of this condition are in agreement with those of Yonelinas (2001, Experiment 1; Yonelinas et al., 1996, Experiment 2), whose subjects received similar instructions.

The neutral condition data are not compatible with the dual-process model. As in Rotello et al. (2004, Experiments 1A and 1B), the subjects provided remember responses to words recognized over a range of confidence levels, and the estimates of R_{DP} obtained from the ROC and from remember responses differed. A possible reaction to these data is to conclude that the neutral instructions encouraged a shift in response strategy that may have contaminated true reporting of remembers and knows. It is important to realize, though, that these instructions are standard, and data produced using them have often been interpreted as if remember and know responses directly reflect separate processes. We will return to this issue in the General Discussion section.

The results of Experiment 1 thus suggest that our two remember–know instructions produced strategies that altered response bias. The role of bias in determining re-
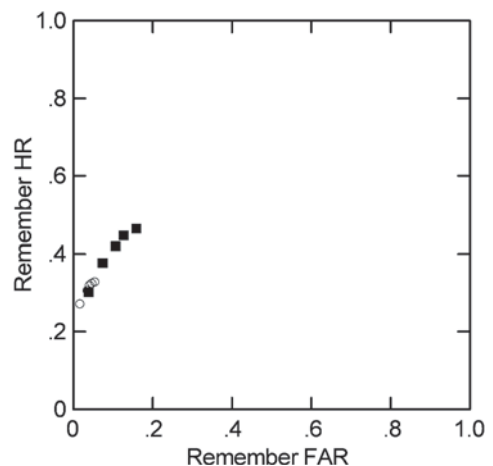

Figure 2. Remember ROC data from Experiment 1. Open circles represent the conservative condition, and closed squares represent the neutral condition. Remember hit rate (HR) = *P*(remember\|old); remember false alarm rate (FAR) = *P*(remember\|new).

member and know responses has been little examined, and data interpretation often implicitly assumes that it does not exist. In Experiment 2, we carefully manipulated remember–know instructions, using the exact instructions in Yonelinas et al. (1996) and Rotello et al. (2004), plus two other instruction sets. In addition, we eliminated the other major difference between the studies by using two study lists. One group of subjects was told to reserve remember responses for cases in which they could report a particular detail to the experimenter. Other groups were shown stimuli with varying amounts of experimenter-provided detail (visual presentation only or visual and auditory presentation) in order to determine the influence of such details on remember responses.

EXPERIMENT 2

Method

Subjects. Fifty undergraduate students at the University of Massachusetts participated for course credit in psychology. All were native English speakers. Data were discarded for 2 subjects who did not follow instructions. The remaining 48 participants were randomly assigned to the *no voices, voices/implicit, voices/explicit*, and *voices/explicit–conservative* conditions (12 in each group).

Stimuli. The 120 words from Experiment 1 were divided into four counterbalanced lists closely matched on the same measures as before. Ten additional words from the pool served as practice, primacy, and recency items. The stimuli were presented on a computer monitor.

Procedure. The experiment consisted of a study phase followed by a practice phase and a final test phase. In the study phase, one of the four word lists (List 1) was presented twice (in the same random order each time); the subjects were informed of the list repetition ahead of time. List 2 was then presented once only (in random order). Each word was presented for 1,200 msec, with a 500-msec interstimulus interval. The subjects were instructed to study all the words carefully in preparation for an upcoming memory test. The remaining word lists were unstudied. In every condition, the lists were counterbalanced such that each served as List 1 for a quarter of the subjects, as List 2 for another quarter, and as an unstudied list for the rest.

The study phase differed across conditions. The subjects in the no-voices condition were simply instructed to study each word in preparation for an upcoming memory test, as in Experiment 1. For the remaining three groups, the presentation of each word was accompanied by a recording of a voice reading it aloud. To replicate the encoding conditions in Yonelinas et al. (1996, Experiment 2), the words in List 1 were read by a male voice on both presentations, and the words in List 2 were read by a female voice. The voices/implicit instructions mentioned that a voice would read each word aloud, but no distinction was made between the male and the female voices, and the subjects were simply instructed to study each word
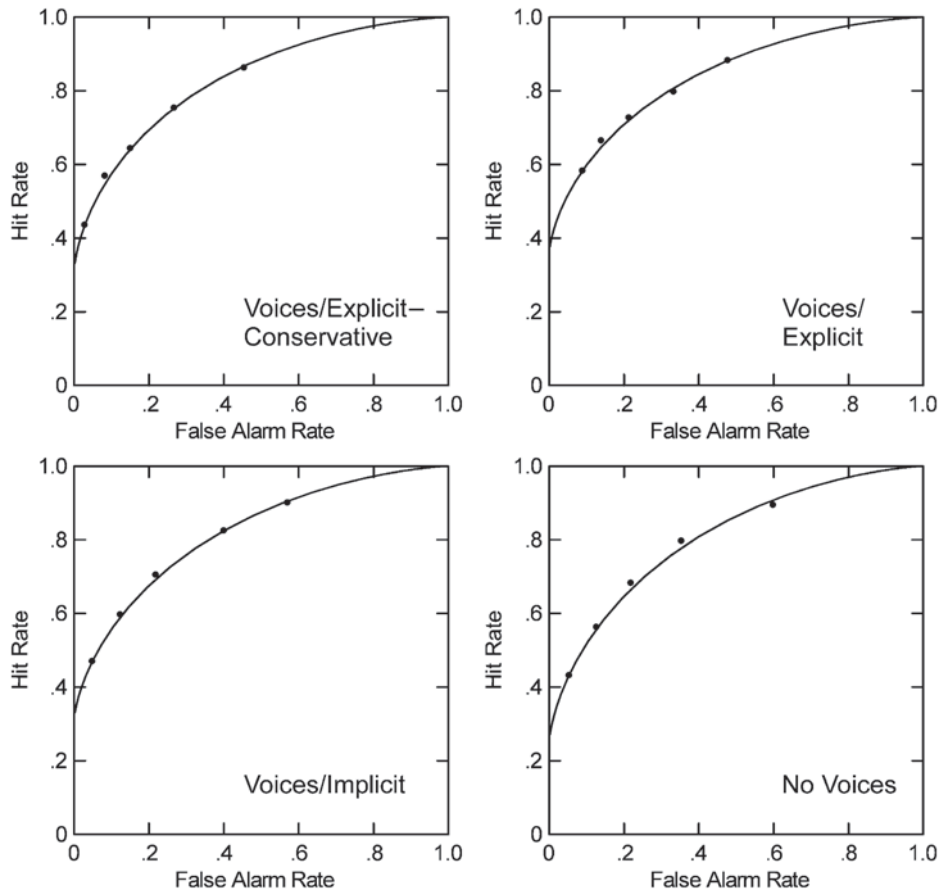

Figure 3. Item recognition ROCs for Experiment 2. The superimposed functions were generated with the dual-process model (Yonelinas, 1994).

for the memory test. The subjects in the voices/explicit and voices/explicit–conservative conditions were explicitly informed about the male voice reading List 1 and the female voice reading List 2. In addition to studying the words, these subjects were instructed to pay particular attention to which voice read each word.

The practice and test phases were similar to those in Experiment 1. The test phase consisted of 60 old and 60 new words presented visually (without voices), a 6-point recognition confidence scale, and a binary remember–know decision. Here, however, the subjects were given different examples of the kinds of specific details that could justify claims of remembering, depending on condition. The no-voices subjects were told that remembering could be based on memory for details, such as the visual appearance of a studied word, its position in the list, or an image or association it cued. In contrast, the subjects in the voices/implicit and voices/explicit conditions were specifically informed that memory for the male or female voice, as well as any of the other kinds of details mentioned to the no-voices group, constituted remembering. Finally, the subjects in the voices/explicit–conservative condition were informed that remembering depended on recalling the male or female voice and that they should not claim to remember a test word unless they were prepared to identify the voice that read it aloud. Although these subjects were asked to report voice during the practice phase, the test phase did not require it. Thus, the voices/explicit–conservative condition most closely replicates the condition in Yonelinas et al. (1996, Experiment 2), and the no-voices condition most closely replicates the neutral condition in Experiment 1.

Results

ROC curves are plotted in Figure 3. (Table A2 in the Appendix contains the confidence-rating data.) The curves overlap substantially but differ in range: The lowest point is obtained in the voices/explicit–conservative condition, and the highest point in the no-voices condition. Curves derived from dual-process theory are fit to the data in Figure 3. The recollection rate R_{DP}, estimated from the y-intercept, ranged from .36 to .25 across conditions, averaging .31 (see Table 1).

Overall response proportions are shown in Table 3. The subjects were more willing to identify (twice-studied) words from List 1 as *old* than words from List 2 and were also more willing to respond that they remembered List 1 words, especially in the no-voices condition. The overall remember hit rate varied across conditions and was lowest in the voices/explicit–conservative condition. Apparently, our instructions to that group to respond *remember* only when they could identify the voice of the speaker effectively reduced remember responses to old items. The estimates of R_{DP} obtained from the overall remember rates averaged .42 and were in poor agreement with the over-

all intercept values, which averaged .31 (see Table 1). The two values agreed best for the voices/explicit–conservative condition (recall that a similar agreement was found for the conservative condition in Experiment 1). For the other three conditions, the average remember rate was .44, whereas the average ROC intercept was .31.

Separate repeated measures 2 (list, within subjects) × 4 (condition, between subjects) ANOVAs on the hit rates and $P(R|\text{old})$ confirmed these observations: The hit rate was significantly affected by study list [$F(1,44) = 52.10$, $MS_e = 0.0051$, $p < .0001$], as was $P(R|\text{old})$ [$F(1,44) = 87.48, MS_e = 0.0116, p < .0001$]. In addition, $P(R|\text{old})$ was influenced by both encoding condition [$F(3,44) = 2.75$, $MS_e = 0.0596$, $p = .054$] and its interaction with study list [$F(3,44) = 2.74, MS_e = 0.0116, p = .055$]. There were no reliable effects of any variable on either the false alarm rate [$F(3,44) = 1.34$, $p = .27$] or $P(R|\text{new})$ [$F(3,44) = 2.04$, $p = .1$].

Proportions of remember responses across old–new confidence levels in the four conditions are plotted as remember ROCs in Figure 4. The remember responses in the voices/explicit–conservative condition were given primarily to words recognized with high confidence, as can be seen by the tight clustering of points in the remember ROC; lower confidence points on the ROC did not include many additional remember responses. This pattern of remember responding agrees with that observed by Yonelinas (2001; Yonelinas et al., 1996) and with that in the conservative condition in Experiment 1. In contrast, the remaining three conditions all show that *remember* responses were given to words recognized with a range of different confidence levels, replicating the pattern of responding observed in the neutral condition in Experiment 1 and in Rotello et al. (2004, Experiment 1).

Discussion

Contrary to dual-process predictions, the major finding of Experiment 2 is that the remember rate does not converge with the ROC intercept. Instead, the remember rate varies across conditions and corresponds to the intercept only in the voices/explicit–conservative condition.

The observation in Yonelinas's experiments that remember responses almost exclusively followed highest confidence *old* judgments has been taken as support for a high-threshold recollective process. Indeed, Yonelinas (2001, 2002; Yonelinas et al., 1996) has equated re-

Table 3
Response Proportions (With Standard Deviations) From Experiment 2

	List 1 Old Items						List 2 Old Items						New Items					
	H		$P(R|\text{Old})$		$P(R|H)$		H		$P(R|\text{Old})$		$P(R|H)$		F		$P(R|\text{New})$		$P(R|F)$	
Condition	M	SD	M	SD	M	SD	M	SD	M	SD	M	SD	M	SD	M	SD	M	SD
Voices/explicit–conservative	.93	.07	.50	.21	.54	.22	.79	.15	.35	.23	.44	.28	.46	.20	.04	.03	.09	.09
Voices/explicit	.91	.09	.70	.18	.77	.17	.85	.09	.53	.18	.62	.19	.48	.23	.13	.14	.27	.22
Voices/implicit	.94	.07	.65	.16	.69	.17	.85	.18	.46	.11	.54	.22	.57	.22	.13	.11	.23	.15
No voices	.96	.05	.66	.20	.69	.21	.82	.12	.35	.16	.43	.20	.60	.18	.11	.13	.18	.21

Note—H, hit rate; $P(R|\text{old})$, $P(\text{remember}|\text{old})$; $P(R|H)$, $P(\text{remember}|\text{hit})$; F, false alarm rate; $P(R|\text{new})$, $P(\text{remember}|\text{new})$; $P(R|F)$, $P(\text{remember}|\text{false alarm})$. H and F are based on rating data with $1 = new$, $2–5 = old$.
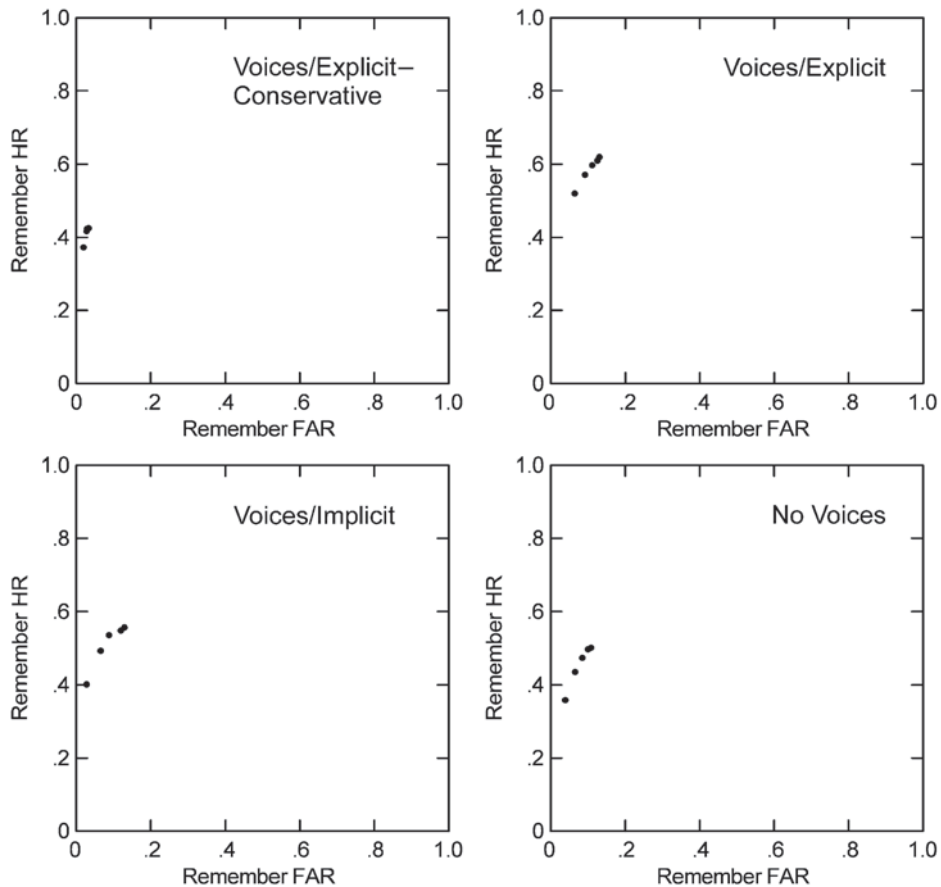
Figure 4. Remember ROC data from the four conditions in Experiment 2. Remember hit rate (HR) = P(remember|old); remember false alarm rate (FAR) = P(remember|new).

member responses with the recollective process in his dual-process model of recognition. In a high-threshold recollection process, only old items can be recollected. According to this view, if remember responses are based on recollection, there should not be many remember responses to new items; those that do occur must be based on nonmemorial noise and, thus, should not vary with *old–new* confidence. The pattern of remember responses reported by Yonelinas and colleagues has been entirely consistent with a high-threshold process.

In the present data set, only the voices/explicit–conservative condition is consistent with this account. The entire pattern of results forces the conclusion that willingness to respond *remember* is strongly influenced by test instructions. The remember response is subject to response bias and does not have an all-or-none character.

GENERAL DISCUSSION

In Experiment 1, standard remember–know instructions led to greater use of the remember response, for both old and new items, than did instructions that placed restrictions on using that response. In Experiment 2, this was true for the comparison between the most conservative condition and the other three. When remember–know

bias is conservative (both experiments) or when a single detail must be recollected (Experiment 2), few remember responses are given, and the resulting remember ROCs approach single-point functions. These data also indicate that remember judgments cannot be pure measures of the underlying recollective process (as has been assumed by many researchers, including Rajaram, 1993, and Gardiner & Richardson-Klavehn, 2000), because response bias affects the subjective reports.

Other studies have also shown systematic changes in the remember false alarm rate. For example, Higham and Vokey (2004) showed that longer identity primes (45 vs. 15 msec) resulted in more remember responses to both old and new target words. Similarly, Hicks and Marsh (1999) found that the remember false alarm rate increased when subjects made remember–know–*new* decisions, instead of *old–new* judgments followed by remember–know responses, and Hirshman and Henzler (1998) reported higher remember false alarm rates when subjects believed more items on the test list had been studied. All of these effects can be understood as changes in response bias along a continuous dimension of memory strength.

An advocate of the dual-process model might argue that these data do not point to a flaw in the model but, instead, show that subjects must be carefully instructed in order for

the remember–know procedure to be valid. That standard instructions give results that contradict the high-threshold model may just mean that many previous experiments have been done poorly.[3]

Alternative interpretations of conflicting empirical results can be handled by two broad strategies. In our view, the best way to understand an experimental design is to construct a model for it. When the basic model fits the data for only a particular variant of the experiment, the ideal solution is to show that *all* the data can be accounted for by changing a parameter of the model. We have done that here by showing that threshold-appearing data reflect a response bias process that applies to all of our experimental conditions.

It is also an option to declare some experimental variants illegal. In the remember–know task, we do not think that would be successful. The standard instructions have been used in many studies (the large majority of the 373 experiments from 63 publications, cited in Rotello et al., 2004). These instructions capture the intuitive distinction between recollection and familiarity, and experiments in which they have been used have been fodder for an extended body of coherent work.

Overall, these data are hard to reconcile with the view that remember judgments measure the same recollective process, as the dual-process or the process-pure views assume. Two continuous models, both based entirely on signal detection processes, appear to provide at least a qualitative account of the data. A one-dimensional model (Donaldson, 1996; Hirshman & Master, 1997) divides a single strength continuum into regions that correspond to the remember, know, and *new* responses. Dunn (2004) and Wixted and Stretch (2004) have shown that this model can account for many features of remember–know data. A two-dimensional model, STREAK (Rotello et al., 2004), hypothesizes two orthogonal criteria in a space defined by global and specific information about the test probes. Both models permit manipulation of criteria to adjust the preponderance of *old* versus *new* and of remember versus know responses and can describe either of the data patterns seen here.

The conclusion that *old–new* judgments are based on a continuum is supported by a long history of ROC curves collected in recognition experiments. Such curves have generally been consistent with a Gaussian, unequal-variance signal detection theory model, as are *old–new* ROCs for the present data.[4] Remember–know judgments have been less studied, but data from the present experiments reveal that they have a similar character. When asked to provide *old–new* ratings, subjects adjust their pattern of responding in accordance with the remember–know instructions so that more or fewer categories are used for the remember response.

CONCLUSION

Remember judgments do not directly and exclusively measure the contribution of a threshold-based recollective process to recognition. Instead, such judgments reflect aspects of continuous memory strength variables and are influenced by subjects' willingness to respond *remember*.

The remember–know paradigm has been extensively used to explore for dissociations, experimental manipulations that affect the rates of remember and know responding differently. The presence of remember–know response bias clouds the interpretation of these dissociation experiments: Changes in response rates could be due to sensitivity changes (the tacit assumption) or to response bias. Signal detection models allow us to distinguish these alternatives.

REFERENCES

COLTHEART, M. (1981). The MRC Psycholinguistic Database. *Quarterly Journal of Experimental Psychology*, **33A**, 497-505.

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

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

GARDINER, J. M., & RICHARDSON-KLAVEHN, A. (2000). Remembering and knowing. In E. Tulving & F. I. M. Craik (Eds.), *The Oxford handbook of memory* (pp. 229-244). New York: Oxford University Press.

HICKS, J. L., & MARSH, R. L. (1999). Remember–know judgments can depend on how memory is tested. *Psychonomic Bulletin & Review*, **6**, 117-122.

HIGHAM, P. A., & VOKEY, J. R. (2004). Illusory recollection and dual-process models of recognition memory. *Quarterly Journal of Experimental Psychology,* **57A**, 714-744.

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.

KUČERA, H., & FRANCIS, W. N. (1967). *Computational analysis of present-day American English.* Providence, RI: Brown University Press.

RAJARAM, S. (1993). Remembering and knowing: Two means of access to the personal past. *Memory & Cognition*, **21**, 89-102.

ROTELLO, C. M., MACMILLAN, N. A., & REEDER, J. A. (2004). Sum–difference theory of remembering and knowing: A two-dimensional signal detection model. *Psychological Review*, **111**, 588-616.

TULVING, E. (1985). Memory and consciousness. *Canadian Psychology*, **26**, 1-12.

WIXTED, J. T., & STRETCH, V. (2004). In defense of the signal detection interpretation of remember/know judgments. *Psychonomic Bulletin & Review*, **11**, 616-641.

YONELINAS, A. P. (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. (2001). Consciousness, control, and confidence: The 3 Cs of recognition memory. *Journal of Experimental Psychology: General*, **130**, 361-379.

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

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.

YONELINAS, A. P., KROLL, N. E. A., DOBBINS, I., LAZZARA, M., & KNIGHT, R. T. (1998). Recollection and familiarity deficits in amnesia: Convergence of remember–know, process dissociation, and receiver operating characteristic data. *Neuropsychology*, **12**, 323-339.

NOTES

1. Yonelinas, Kroll, Dobbins, Lazzara, and Knight (1998) list three estimates of recollection based on the remember hit rate (RH) and the remember false alarm rate (RF): RH itself, $RH - RF$, and $(RH - RF)/(1 - RF)$. The estimate most commonly used (and used here) is $RH - RF$, which is always the smallest of these.

2. Three other experiments collected both ratings and remember–know judgments (Yonelinas, 2001, Experiments 2A–2C). However, in

those experiments, remember responses were assumed to be highest confidence *old* decisions, so the estimates of R_{DP} from the ROC and the remember–know paradigms are confounded.

3. One might argue that only the studies with low remember false alarm rates—.05 or less, say—should be considered valid, because only in those studies were the instructions properly understood. Using such a criterion would arbitrarily exclude one third of the remember–know conditions in the literature (mean remember false alarm rate = .053, SD = .045, maximum = .24 in the database of 400 cases reported by Dunn, 2004), including some, but not all, conditions within particular experiments. For example, it would suggest that in Hirshman and Henzler's (1998) study, subjects who were told that 70% of the test items had been studied did not understand the remember–know distinction but that those who were told that 30% of the items were old *did* make the distinction appropriately.

4. Although we fit the dual-process model to the ROCs in this study, all of the fits are statistically equivalent to those of an unequal-variance signal detection model.

APPENDIX
Frequency Data From Each Experiment: Old–New Confidence Ratings and Binary Remember–Know Judgments

Table A1
Experiment 1

Items	Rating	Conservative Condition			Neutral Condition		
		Remember	Know	New	Remember	Know	New
Old	6	388	100	–	433	59	–
	5	49	82	–	109	92	–
	4	19	104	–	61	101	–
	3	8	126	–	38	106	–
	2	5	160	–	26	122	–
	1	–	–	394	–	–	290
New	6	25	30	–	55	13	–
	5	26	42	–	51	45	–
	4	9	65	–	47	84	–
	3	8	150	–	29	124	–
	2	11	200	–	46	199	–
	1	–	–	870	–	–	745

Note—Ratings: 6 = *sure studied*, 2 = *maybe studied*, 1 = *sure new*.

Table A2
Experiment 2

Items	Rating	Voices/ Explicit– Conservative			Voices/ Explicit			Voices/ Implicit			No Voices		
		R	K	N	R	K	N	R	K	N	R	K	N
List 1													
Old	6	160	31	–	217	27	–	173	23	–	175	29	–
	5	15	35	–	23	8	–	38	13	–	35	21	–
	4	4	21	–	8	10	–	17	20	–	15	23	–
	3	0	38	–	4	11	–	5	31	–	10	19	–
	2	0	31	–	1	19	–	2	19	–	1	21	–
	1	–	–	25	–	–	31	–	–	19	–	–	11
List 2													
Old	6	107	14	–	156	18	–	115	27	–	82	24	–
	5	17	29	–	14	14	–	28	12	–	20	18	–
	4	1	28	–	11	15	–	14	27	–	13	35	–
	3	0	41	–	5	31	–	4	46	–	7	46	–
	2	1	46	–	6	35	–	4	30	–	2	47	–
	1	–	–	74	–	–	54	–	–	53	–	–	65
Combined Lists													
Old	6	267	45	–	373	45	–	288	50	–	257	53	–
	5	32	64	–	37	22	–	66	25	–	55	39	–
	4	5	49	–	19	25	–	31	47	–	28	58	–
	3	0	79	–	9	42	–	9	77	–	17	65	–
	2	1	77	–	7	54	–	6	49	–	3	68	–
	1	–	–	99	–	–	85	–	–	72	–	–	76
New Item List													
New	6	15	6	–	47	17	–	21	14	–	29	9	–
	5	6	32	–	20	16	–	27	26	–	19	33	–
	4	1	48	–	14	39	–	16	52	–	14	51	–
	3	1	82	–	10	75	–	23	106	–	11	85	–
	2	2	132	–	4	99	–	7	114	–	6	168	–
	1	–	–	389	–	–	373	–	–	305	–	–	284

Note—R, remember response; K, know response; N, new judgment (confidence rating = 1). Ratings: 6 = *sure studied*, 2 = *maybe studied*, 1 = *sure new*.