Analysis of RT distributions in the remember–know paradigm

CAREN M. ROTELLO AND MIN ZENG

University of Massachusetts, Amherst, Massachusetts

Do remembering and knowing differ qualitatively (reflecting distinct underlying processes) or quantitatively (reflecting different levels of strength)? Broadly speaking, models of remember–know judgments based on these alternatives have been tested by examining the proportion of remember and know responses that are made across conditions or levels of confidence. Here, we consider reaction time (RT) data. We replicate Dewhurst and Conway's (1994) observation that *old* judgments followed by remember responses are faster, on average, than those followed by know decisions, but show that this effect is largely due to differing distributions of remember and know responses across confidence levels. In addition, fits of ex-Gaussian distributions of hit RTs followed by either remember or know judgments yield similar parameter values when confidence level is controlled. Thus, these RT data do not provide strong support for the idea that remembering and knowing depend on different processes.

The remember–know paradigm (Tulving, 1985) is used to assess the contributions of recollection and familiarity to recognition judgments. In this task, subjects make oldnew recognition decisions on both studied items and lures. For each stimulus judged old, subjects report whether they remember something specific about studying the item or feel as if they know that they studied it, despite its failure to evoke conscious recollection of any details. Systematic differences in remember-know response rates for various empirical manipulations suggest that the responses depend on different underlying processes (Gardiner & Richardson-Klavehn, 2000). However, the data have also been interpreted in terms of a one-dimensional model in which remember responses are made to stronger memories than are know judgments (Donaldson, 1996; Dunn, 2004; Wixted & Stretch, 2004).

Reaction time (RT) data provide another window onto underlying processes, yet few studies have reported RTs in this literature. In the earliest report, Dewhurst and Conway (1994) found that *old* decisions that were followed by a remember response were 500–700 msec faster than those followed by know judgments. This pattern has been replicated in all the studies that have analyzed RTs (e.g., Dewhurst, Holmes, Brandt, & Dean, 2006; Wixted & Stretch, 2004); it has usually been interpreted as reflecting the use of different underlying processes (cf. Wixted & Stretch, 2004). One interpretation of the data is that know responses are slow because they are made only if remembering fails (Murdock, 2006).

Before accepting these differences in mean RT as reflecting distinct underlying processes, a confound in the data must be considered. Specifically, RTs are negatively correlated with decision confidence (Petrusic & Baran-

ski, 2003), and remembers are generally associated with greater confidence than are knows (Rotello, Macmillan, & Reeder, 2004). In our experiments, we attempted to control confidence level when evaluating *old–new* RTs followed by remember and know judgments.

A second goal of the present experiments was to evaluate the characteristics of the RTs more thoroughly, using a distributional analysis. Mean RT can differ because of changes in the location or shape (i.e., skew) of the distribution of responses; these effects have different psychological interpretations. Location changes, which occur when all responses are slowed down by a constant amount in one condition relative to another, are often attributed to differences in the contribution of automatic (*residual*) processing to the judgments (Balota & Spieler, 1999; Hockley, 1984). In contrast, shape changes, which occur when responses are slowed by a variable amount, have been interpreted as reflecting differences in strategic (*decision*) processing (Heathcote, Popiel, & Mewhort, 1991; Ratcliff & Murdock, 1976; Rouder, Lu, Speckman, Sun, & Jiang, 2005).

Information about whether remember responses are faster than know responses because of location or shape changes, or both, would constrain model development. To this end, we collected large amounts of data for each subject and fit the resulting RT distributions with ex-Gaussian functions (Balota & Spieler, 1999; Hockley, 1984; Hohle, 1965). The ex-Gaussian is a convolution of normal and exponential distributions; it has three parameters: μ and σ , the normal's mean and standard deviation, and τ , the mean of the exponential. Rouder et al. (2005) showed that μ and σ reflect the location and scale of the ex-Gaussian, and the ratio τ/σ reflects its shape. The skew of the ex-Gaussian is $2\tau^3$, which has perhaps led to the focus on τ alone as a mea-

C. M. Rotello, caren@psych.umass.edu

sure of decision processing (Heathcote et al., 1991). We follow suit, evaluating τ as a measure of whether remember and know responses reflect the same decision processes.

EXPERIMENT 1

Experiment 1 was a list length experiment using the remember–know paradigm with a large number of test trials so that distributional analyses could be conducted for individual subjects. On each test trial, the subjects made an *old–new* judgment, rated their confidence that the item was old, and provided a remember–know response.

Method

Subjects. Four members of the psychology department community at the University of Massachusetts at Amherst participated in exchange for a \$40 cash payment.

Materials and Design. A pool of 3,072 nouns was created; the words varied widely in written frequency (10–4,500; Kučera & Francis, 1967) and length (3–15 letters). They were randomly assigned to roles. (1) Study words on short (30-word) or long (90-word) lists; 720 words were assigned to each list length, yielding 24 short and 8 long study lists. (2) An equal number of words served as lures on the recognition test that followed each study list. (3) Finally, 3 untested words appeared at the beginning and end of each list to absorb primacy and recency effects.

The subjects participated individually in eight experimental sessions. Each session consisted of three short and one long study-test blocks, in a random order. Following each study list, there was a recognition test on all of the studied words from that block, plus an equal number of lures, in a random order.

Procedure. The subjects studied the words in each list in preparation for a recognition test. Each word was shown for 1,750 msec, with a 250 msec interstimulus interval. A recognition test immediately followed each list; the subjects were informed that only lures and words from the immediately preceding study list would appear on each test. Standard remember–know instructions (Rajaram, 1996) were used, and the subjects were encouraged to take breaks between blocks.

For each test word, the subjects made three judgments; each was cued visually in turn. First, they made an old-new recognition decision using their middle fingers, pressing buttons 2 (old, left hand) or 5 (new, right hand) on a six-button Cedrus response box (RB-610) with 1-msec time resolution. The RTs to respond old are the primary data of interest. Second, they rated their old-new confidence $(1 = sure\ new, 6 = sure\ old)$ on the response box, using three fingers of each hand; use of the full scale allowed the subjects to reverse their judgments (Van Zandt & Maldonado-Molina, 2004). Finally, the subjects made remember-know-new judgments by using their ring fingers to press either 1 (know or new) or 6 (remember); new and know responses were later distinguished by their confidence rating (assuming that know judgments followed ratings of 4-6 and new followed 1-3).2 The subjects were asked to make their *old-new* decisions as quickly as possible without sacrificing accuracy but were not under time pressure for the other judgments. The next word appeared 750 msec after the rememberknow-new judgment.

Results

We examined three aspects of the data: (1) memory accuracy and remember–know response proportions; (2) mean RT for hits followed by remember or know responses; and (3) RT distributions for hits that were followed by remember or know responses, fit with ex-Gaussian distributions.

Response proportions. All the subjects made more accurate recognition judgments on the short lists (mean

 d_a : short = 1.75, long = 1.18).³ Table 1 shows the probability that the subjects responded *remember* or *know*: Longer lists resulted in fewer remember responses; know rates were not affected (Cary & Reder, 2003).

Mean RTs to *old* **judgments**. The mean RTs for *old* responses followed by remember and know responses are shown in Table 2. The overall hit RTs, collapsed across confidence level, replicate the results in the literature in showing faster *old* responses followed by remember than by know responses (all 4 subjects showed the effect: minimum $t_{\text{observed}} = 7.2$). On average, the RT advantage for remember hits was 187 msec for short lists and 183 msec for long lists.

All the subjects also showed significantly shorter RTs with greater confidence (long condition mean r = -.90; short mean r = -.95). Considering only the highest confidence hits, the difference in mean RTs for remember and know responses is reduced (short, 24 msec; long, 97 msec). Similarly, hits made with a confidence of 5 also showed a reduced (or reversed) advantage for remember judgments (short, 118 msec; long, −56 msec). At a confidence level of 4, there were too few remember hits for meaningful comparison. After a Bonferroni correction for multiple comparisons, only 1 subject (C, long condition, confidence of 5) showed a significant, and reversed, effect of rememberknow status on confidence-controlled hit RT [remember = 1,709, know = 1,008; t(33) = 3.68, corrected p < .05]. However, these null results may be partly due to lack of power: The sample sizes are not always large. For false alarms, the remember responses are slower than the know responses, but small samples limit our conclusions.

RT distributions. We compared *old* response RT distributions at a particular confidence level that was selected so as to maximize the available data for each subject. Subjects B and C both spread their remember and know responses across confidence levels, but neither of them provided enough responses of both types at the same confidence rating that both RT distributions could be fit (the maximum number of responses in the smaller category at a particular confidence level was 10 for Subject B, 14 for Subject C). Therefore, we will focus on Subjects A (confidence level of 5) and D (confidence of 6). Similarly, we will concentrate on hit rather than on false alarm RTs.

Figure 1 shows the RT distributions for the hits that were assigned remember and know judgments in the long-(upper half) and short- (lower half) list conditions, separately for Subjects A and D. Superimposed on each empirical distribution is the best-fitting ex-Gaussian function

Table 1
Mean Probabilities of Remember (R) and Know (K) Responses for All Subjects in Both Experiments With Standard Errors

		Response Proportion										
	List Length		Old	Items		New Items						
			3	ŀ	ζ	I	3	K				
		\overline{M}	SE	\overline{M}	SE	\overline{M}	SE	\overline{M}	SE			
Experiment 1	Short	.51	.10	.31	.10	.02	.01	.12	.03			
•	Long	.40	.12	.31	.10	.03	.01	.17	.05			
Experiment 2		.37	.09	.36	.11	.06	.05	.19	.05			

							0									
			Item Type													
			Old						New							
		R	Remember			Know			Remember			Know				
	List	R	Γ		RT	Γ		R	Γ		R	Γ				
Confidence Level	Length	M	SE	n	\overline{M}	SE	n	M	SE	n	M	SE	n			
All old responses	Short	833	24	1,433	1,020	15	860	1,191	68	40	1,276	30	333			
	Long	913	12	1,140	1,096	15	860	1,427	119	76	1,284	27	470			
6	Short	828	25	1,356	852	19	187	1,054	65	28	1,074	112	3			
	Long	886	11	1,061	983	25	136	1,295	170	50	1,035	198	9			
5	Short	909	23	70	1,027	31	273	1,514	173	7	1,212	67	64			
	Long	1,156	58	62	1,100	26	263	1,630	146	11	1,402	65	96			
4	Short	1,051	152	7	1,094	22	400	1,503	216	5	1,293	33	266			
	Long	1,657	180	17	1,128	22	461	1,720	165	15	1,259	29	365			

Table 2
Mean *Old* Response Times (RTs, in Milliseconds; With Standard Errors) and Sample Sizes (ns), for Judgments Followed by Remember and Know Responses at Each List Length and Level of Confidence in Experiment 1

obtained using quantile maximum probability estimation (QMPE; Heathcote, Brown, & Cousineau, 2004); the parameter values are shown.⁵ The estimated values of τ are somewhat smaller for the short lists than for the long lists, a sensible finding considering that skew tends to increase with mean RT (Ratcliff, 1993).

The key data are the parameter values for the old responses followed by remember and know judgments. For each subject, the values of τ are similar within list length. To determine whether the differences in τ were statistically significant, we estimated 95% confidence intervals (CIs) for the difference ($\tau_R - \tau_K$), as well as differences in μ and σ , using a nonparametric bootstrapping technique (Efron & Tibshirani, 1993). For each subject, we sampled $N_{\rm R}$ and $N_{\rm K}$ hit RTs from the observed set, with replacement, where N_i equaled the number of hit RTs followed by remember or know at the selected confidence level (e.g., for Subject A's short condition, $N_R = 53$ and $N_K =$ 113). This sample of RTs was then fit with ex-Gaussian distributions using QMPE, exactly as for the original data. From the resulting parameter values, three differences were obtained: $\tau_R - \tau_K$, $\mu_R - \mu_K$, and $\sigma_R - \sigma_K$. This sampling process was repeated a total of 2,000 times for each subject, and the 95% CI for each difference was inferred; differences at the 2.5 and 97.5 percentiles were chosen as the upper and lower bounds. The resulting estimated CIs are shown in Table 3; all the CIs include zero, so we cannot reject the null hypothesis that the individual parameter values of the ex-Gaussian distributions are the same for confidence-matched hits followed by remember or know judgments.

Discussion

We draw two primary conclusions: Mean RTs to *old* responses that are followed by remember responses are indeed faster than those followed by know responses (Dewhurst & Conway, 1994), but that difference is partly attributable to differences in the distribution of remember and know responses over confidence ratings. When confidence level is controlled, the RT differences across remember and know judgments are greatly reduced. In addition, the individual parameters of the ex-Gaussian distributions do not differ

significantly for confidence-controlled remembers and knows. Before we consider the theoretical implications of these data, we first will replicate the basic effects with additional subjects. In addition, we will modify an aspect of the procedure that may have encouraged the subjects to make *sure old* and remember responses together: The same response key was used for both judgments.

EXPERIMENT 2

The primary purpose of Experiment 2 was to replicate the RT analyses of Experiment 1, with more data per subject and without biasing the subjects to respond *remember* after *sure old* judgments. Elimination of the list length factor allowed us to double the number of observations per cell without increasing the number of study—test cycles. In addition, we changed the way confidence responses were made, to reduce the strong pairing of *sure old* and remember responses: The subjects used a mouse to select a location on a continuous confidence bar and pressed a key to indicate remembering. Use of the confidence bar allowed us to bin the RT data so as to maximize the number of *old* responses followed by remember and know responses at a particular range of confidence levels.

Method

Subjects. Four members of the psychology department community at the University of Massachusetts at Amherst participated in exchange for an \$80 cash payment. None had participated in Experiment 1.

Materials and Design. The stimuli were the same as those in Experiment 1, but all the study lists consisted of 3 primacy words, 90 critical words, and 3 recency words. The subjects participated individually in six experimental sessions consisting of two study–test blocks each.

Procedure. The procedure was the same as that in Experiment 1, except for changes to the equipment and confidence ratings. We used an eight-button Cedrus response box (RB-830). The subjects made *old–new* judgments on the response box, using their thumbs (left = old; right = new). Then they were presented with a vertical bar, 468 pixels long, on the right half of the computer screen; it was labeled "sure" on the top and "guess" on the bottom. The subjects used a computer mouse to select a confidence level from this scale. Finally, the subjects made a trinary remember–know–*new* judgment on the response box (left index finger = remember; right index finger = know; right thumb = new). All of the response keys were marked for the subjects.

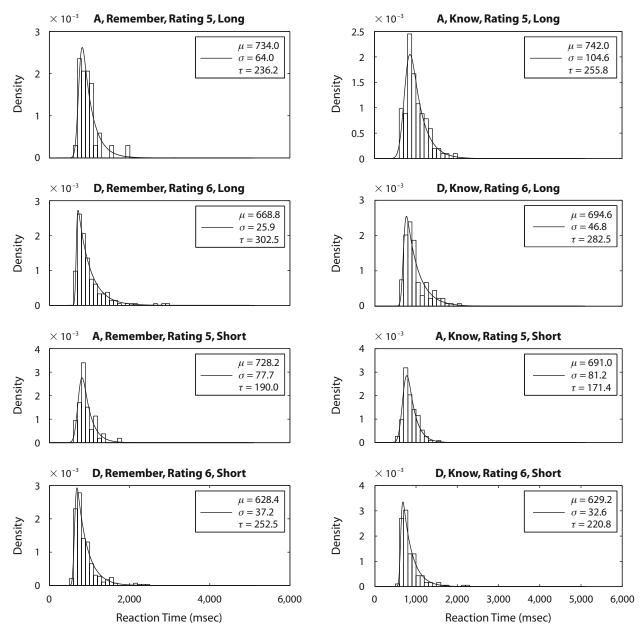


Figure 1. Reaction time distributions for hits followed by remember (left column) or know (right column) judgments in the long-list (upper half) and short-list (lower half) conditions, separately for Subjects A and D in Experiment 1. Subject A's data are shown for a confidence level of 5; Subject D's data are shown for a confidence level of 6. In all the panels, the empirical data are shown as a histogram, and the best-fitting ex-Gaussian distribution is superimposed with a solid contour. The parameters of the ex-Gaussian are shown.

Results

Response accuracy. Performance was similar to that in the long-list condition in Experiment 1, in terms of both accuracy (mean $d_a = 1.25$) and the remember response rate.

Mean RTs to *old* **judgments**. The mean RTs for *old* responses followed by remembers and knows are shown in Table 4 for each subject. Ignoring confidence level, every subject made faster *old* responses when they next responded *remember* rather than *know* (minimum $t_{\text{observed}} = 2.6$), replicating the results of Experiment 1. The difference in hit RT averages was 270 msec.

To evaluate mean RT while controlling for confidence, we used two ways of dividing the continuous confidence ratings into bins. First, we divided the scale into three 156-pixel segments that reflected higher (upper third), moderate (middle third), and lower (lower third) levels of confidence. Table 4 shows that most of the responses fell in the upper third category. At this confidence level, where samples sizes are maximized, all the subjects showed a numerically reduced RT advantage for remember hits, as compared with know hits (mean difference = 94 msec). When the mean RTs for upper third hits were evaluated, however, only Subject A showed a significant RT advantage for re-

Confidence Intervals (6 for <i>Old</i> Responses Followed by Re	,	ection Time Distribution Param es, by Subject and List Length i	
	μ	σ	τ
	95% CI	95% CI	95% (

Table 3

				μ				σ					
Subject/					95% CI			95% CI			95% CI		
Confidence	List	Sample	Size		Lower	Upper		Lower	Upper		Lower	Upper	
Level	Length	Remember	Know	$\mu_{\rm R} - \mu_{\rm K}$	Bound	Bound	$\sigma_{\rm R} - \sigma_{\rm K}$	Bound	Bound	$\tau_{\mathrm{R}} - \tau_{\mathrm{K}}$	Bound	Bound	
A/5	Long	34	102	-8.0	-92	139.7	-40.5	-104.8	81.6	-19.6	-185.5	108.6	
	Short	53	113	37.2	-56.1	112.5	-3.5	-79.1	55.3	18.6	-71.1	128.9	
D/6	Long	214	134	-25.8	-63.1	4.6	-21.0	-52.5	10.5	20.0	-55.3	95.1	
	Short	291	185	-0.8	-25.8	20.4	4.6	-11.8	22.9	31.7	-24.0	87.1	

Note—Data are for a particular confidence level for each subject; CIs were estimated with a nonparametric bootstrap (see the text).

member responses over know responses. Second, we chose a narrower segment (the *top slice*) of the confidence bar that maximized the overlap of remember and know responses. This analysis eliminated confidence levels that were followed primarily by know judgments; the top slice was determined for each subject individually and ranged in size from 20 to 60 pixels. Using this approach, 2 subjects (A and B) continued to show numerically faster remember hits and 2 (C and D) showed numerically slower remember hits; only Subject C's RT difference was significant.

None of the subjects' mean false alarm RTs differed for remember or know responses, although sample sizes limit our conclusions. In addition, none of the subjects made remember false alarms more quickly than know hits.

RT distributions. The primary data of interest are the distributions for hit RTs followed by remember and know responses, with *old–new* confidence level controlled. Our use of the confidence bar allowed selection of confidence levels at which both remember and know responses were

represented in large numbers; every subject's data were subjected to the ex-Gaussian analysis.

Figure 2 shows the observed RT distributions for the hits made with top slice confidence that were followed by remembers and knows, separately for each subject. The bestfitting ex-Gaussian distribution is superimposed on each empirical distribution, and the parameter values are shown. With the exception of Subject C, whose top slice sample size was relatively small, the τ parameters are quite similar for remember and know judgments for each subject. CIs for the differences in τ , as well as μ and σ , were obtained using the bootstrap method described for Experiment 1; the results are shown in Table 5. Consistent with the data from Experiment 1, the CIs for $\tau_R - \tau_K$, $\mu_R - \mu_K$, and $\sigma_R - \tau_K$ $\sigma_{\rm K}$ include zero for all the subjects. Thus, when confidence is approximately matched, we cannot reject the null hypothesis that the individual parameters of the ex-Gaussian distribution for *old* responses followed by remember equal those followed by know. When confidence is more loosely

Table 4
Mean *Old* Response Times (RTs, in Milliseconds; With Standard Errors) and Sample Sizes (ns) for Judgments Followed by Remember and Know Responses, by Confidence Level in Experiment 2

	Item Type													
				О	ld	New								
		Re	Remember			Know			Remember			Know		
		R	RT		RT			RT			RT			
Subject	Confidence Level	M	SE	n	M	SE	n	M	SE	n	M	SE	n	
A	All old responses*	1,089	27	673	1,266	47	273	1,309	52	228	1,425	68	152	
	Top slice (20 pixels)	1,046	34	372	1,133	80	91	940	71	16	1,390	418	5	
	Upper third (156 pixels)*	1,075	29	568	1,234	51	223	1,333	75	110	1,469	109	61	
	Middle third	1,229	99	72	1,539	164	31	1,274	109	58	1,490	126	54	
	Lower third	1,006	54	33	1,194	140	19	1,298	95	60	1,257	112	37	
В	All old responses*	1,018	19	282	1,083	16	400	1,112	123	7	1,180	23	211	
	Top slice (60 pixels)	1,037	21	269	1,085	43	107	1,236	136	5	1,163	102	11	
	Upper third (156 pixels)	1,015	19	277	1,044	20	262	1,166	131	6	1,112	27	92	
	Middle third	1,157	231	5	1,150	31	97	787	_	1	1,184	43	75	
	Lower third	_	_	0	1,171	54	41	_	_	0	1,313	59	44	
C	All old responses*	1,716	57	367	2,306	88	223	2,801	284	22	3,117	155	109	
	Top slice (40 pixels)*	1,563	56	279	1,316	88	47	1,525	321	3	_	_	0	
	Upper third (156 pixels)	1,628	53	344	1,780	107	110	2,349	315	11	3,622	405	21	
	Middle third	3,022	342	23	2,787	126	108	3,077	453	10	2,996	164	88	
	Lower third	_	_	0	3,455	165	5	5,000	_	1	_	_	0	
D	All old responses*	916	27	258	1,164	27	646	1,439	351	10	1,529	48	341	
	Top slice (20 pixels)	907	29	234	901	38	172	693	118	2	1,183	446	3	
	Upper third (156 pixels)	912	27	254	948	28	335	1,007	167	8	1,359	124	26	
	Middle third	1,016	20	2	1,272	60	104	2,158	_	1	1,239	70	70	
	Lower third	1,277	415	2	1,460	59	207	4,179	_	1	1,629	61	245	

^{*}Hits followed by remember are faster than hits followed by know; Bonferroni-corrected p < .05.

controlled, in the upper third analysis, the parameter values for Subject A's data indicate that the mean RT advantage for remember hits was attributable to the τ parameter (i.e., the know distribution had more skew). In contrast, for Subjects B, C, and D, the upper third analysis yields the same conclusions as the top slice analysis.

Discussion

Two key findings replicated the results of Experiment 1 with an improved design that permitted more comparisons

to be made. Overall, the mean RT for hits followed by remember responses was again found to be significantly shorter than that for hits followed by know responses. However, when confidence was approximately matched, the magnitude of the RT advantage for remembers was reduced. Similarly, we replicated the observation that the individual parameters of the RT distributions for hits followed by remember judgments did not differ from those followed by know responses when confidence was well matched. Although these results agree with those in Ex-

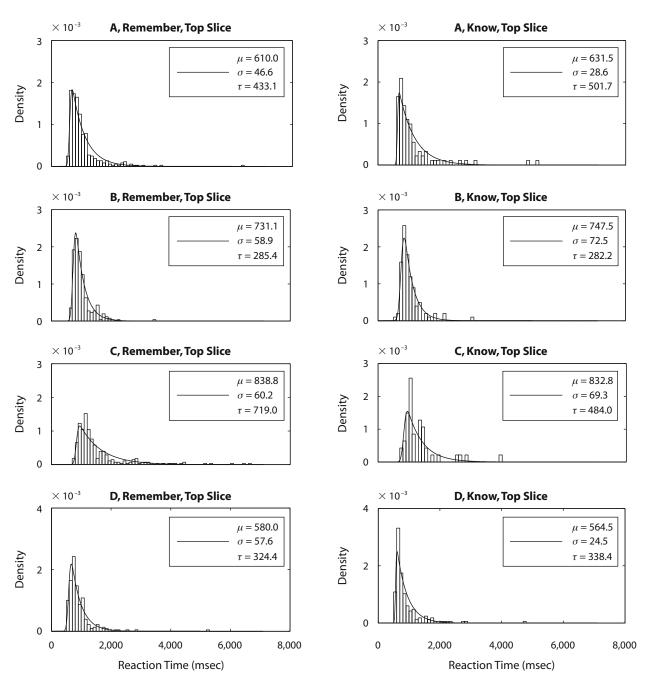


Figure 2. Reaction time distributions for hits followed by remember (left column) or know (right column) judgments, separately for each subject (by rows) in Experiment 2. Data are from the top-slice analysis. In all the panels, the empirical data are shown as a histogram, and the best-fitting ex-Gaussian distribution is superimposed with a solid contour. The parameters of the ex-Gaussian are shown.

				μ			σ		au		
				95% CI			95% CI			95% CI	
	Sample Size			Lower	Upper		Lower	Upper		Lower	Upper
Subject	Remember	Know	$\mu_{\mathrm{R}} - \mu_{\mathrm{K}}$	Bound	Bound	$\sigma_{\rm R} - \sigma_{\rm K}$	Bound	Bound	$ au_{ m R} - au_{ m K}$	Bound	Bound
A	371	91	-21	-77	115	18	-17	60	-69	-380	136
В	256	101	-15	-61	32	-12	-50	44	1	-101	98
C	277	47	-5	-80	70	3	-45	70	231	-22	466
D	231	166	-1	-22	75	7	-8	38	5	-159	92

Table 5
Confidence Intervals (CIs) for the Differences in Reaction Time Distribution Parameters for *Old* Responses Followed by Remember and Know Judgments, by Subject in Experiment 2

Note—Data are from the top slice analysis; CIs were estimated with a nonparametric bootstrap (see the text).

periment 1, the power of our tests is low (averaged .24 for the top slice mean RT analysis, using the observed RT difference as the mean of the alternative distribution), and additional research is needed.

GENERAL DISCUSSION

Our data concur with those in the literature: Mean RTs to respond *old* are faster when subjects subsequently claim to remember details of their prior experience with the stimulus than when they merely know that it was studied (Dewhurst & Conway, 1994). However, the confidence with which subjects respond *old* is also higher, on average, for items judged to be remembered rather than known. Our analyses demonstrate that variation in confidence level drives much of the difference in remember and know RTs. When confidence is roughly equated, the difference in mean hit RTs for remembered and known stimuli is much reduced. Indeed. the means did not differ significantly for the sample sizes tested here. Our distributional analyses also suggest that when confidence levels are similar, we cannot reject the null hypothesis that the individual parameters of the ex-Gaussian distributions are equal for remember and know responses.

One implication of our confidence-controlled distributional analysis is that remember and know responses may be based on the same underlying memory information and processes: Neither the μ parameter of the ex-Gaussian, which reflects automatic processing (Hockley, 1984), nor the τ parameter, which reflects decision processing (Heathcote et al., 1991), differed as a function of the subjective basis of the *old* decision. Thus, our data are broadly consistent with the one-dimensional signal detection model of remember-know judgments, which assumes that remember responses are made to items with greater memory strength and know responses to those with lower strength (Donaldson, 1996). Wixted and Stretch (2004) proposed a version of this model in which the remember-know criterion varies across trials, thus allowing both remember and know responses to be associated with the same memory strength on different trials. If RT is a function of distance from a response criterion (Murdock, 1985), this model accounts for the correlation between confidence and RT (Petrusic & Baranski, 2003), as well as for both of our major effects: faster old responses for remembered items when confidence is ignored and more similar *old* RTs for remembered and known items when confidence is controlled.

We simulated this model by sampling 10,000 target and lure strengths, assuming that targets were distributed as N(1.25, 1.25) and lures as N(0, 1). The old-new decision criterion (C_0) was fixed at 0.625, and the remember–know criterion (C_r) had a variable location (M=1.7, SD=1). These parameters yielded hit and false alarm rates of .69 and .27, and corresponding remember rates of .37 and .09, approximating our data. Each sampled strength (s_i) was compared with C_0 and declared to yield an old response if it exceeded that criterion. Similarly, it was compared with C_r and judged to elicit a remember response if it exceeded that value. The strengths were converted to RTs via Equation 1, which converts distance to criterion to RT, using an arbitrary exponential function (e.g., Murdock, 1985):

$$RT = 700 + 450e^{-0.6(s_i - C_o)}.$$
 (1)

The resulting mean RTs for remember and know hits were 888 and 995 msec, a 107-msec advantage for remember responses that is a bit smaller than that in Table 1. The parameters in Equation 1 were not optimized, so we will focus only on whether the predicted RT difference decreases when confidence is controlled. When lower confidence responses ($s_i < 1.2$) were excluded, the mean RTs were 868 and 931 msec, a smaller (63-msec) advantage for remember responses analogous to that in our data. Results derived from other cutoff values revealed that the remember advantage was inversely related to the cutoff value but that it was never eliminated. These simulated RTs suggest that our failure to observe a significant RT advantage for confidence-controlled remember responses may have been due to low power. Alternatively, the onedimensional model (or the RT distance mapping we assumed) may not capture the processes accurately.

Although the simulated mean RTs are reassuring about the ability of the one-dimensional model to explain broad aspects of the data, more detailed analyses suggest problems. For example, the remember false alarms (ignoring confidence) were not made more quickly than know hits (Wixted & Stretch, 2004), even though those falsely recognized stimuli had greater average strength. More troubling, the simulated RT distributions for known items were left-skewed, contrary to the data. Additional research is needed to assess the generality of these preliminary simulations.

Finally, an alternative theoretical view is that remember and know judgments are based on different underlying processes (Murdock, 2006; Reder et al., 2000; Yonelinas, 2001). These models assume that know responses are made only if remembering fails, consistent with faster old judgments for remembered stimuli. The predictions of these models for confidence-controlled RTs are less obvious, because they do not specify how confidence ratings and subjective reports of remembering and knowing are combined. Yonelinas has assumed that remember judgments necessarily occur at the highest confidence levels and know judgments at lower confidence levels. Given that assumption, our evaluation of RTs at a common confidence level is inappropriate. However, the assumption is counterfactual: The subjects in both experiments provided remember and know responses at overlapping confidence levels. Although these RT data appear to raise problems for models that assume that remembering and knowing differ qualitatively, strong claims cannot be made until the models are better specified.

Many analyses of accuracy data show the one-dimensional model to be at least as viable as the dual-process model in accounting for remember–know data (Dunn, 2004) and recognition data more generally (Wixted, 2007). Our analysis of these RT data provides a completely different type of evidence on the nature of remembering and knowing yet leads to the same conclusion: Remember and know responses appear to be based on the same underlying memory information. Despite this unanimity, the dominant models of remember–know judgments require additional development to fully account for these results.

AUTHOR NOTE

This research was supported by Grant R01 MH60274 from the National Institutes of Health. Experiment 1 was part of M.Z.'s master's thesis. We thank Neil Macmillan for ongoing discussions of remember–know models and for his detailed comments on an earlier version of the manuscript. Bill Hockley and E. J. Wagenmakers provided extremely helpful reviews. Correspondence concerning this article should be addressed to C. M. Rotello, Department of Psychology, Box 37710, University of Massachusetts, Amherst, MA 01003-7710 (e-mail: caren@psych.umass.edu).

REFERENCES

- BALOTA, D. A., & SPIELER, D. H. (1999). Word frequency, repetition, and lexicality effects in word recognition tasks: Beyond measures of central tendency. *Journal of Experimental Psychology: General*, 128, 32-55.
- CARY, M., & REDER, L. M. (2003). A dual-process account of the list-length and strength-based mirror effects in recognition. *Journal of Memory & Language*, 49, 231-248.
- COUSINEAU, D., & LAROCHELLE, S. (1997). PASTIS: A program for curve and distribution analyses. *Behavior Research Methods, Instru*ments, & Computers, 29, 542-548.
- DEWHURST, S. A., & CONWAY, M. A. (1994). Pictures, images, and recollective experience. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **20**, 1088-1098.
- DEWHURST, S. A., HOLMES, S. J., BRANDT, K. R., & DEAN, G. M. (2006). Measuring the speed of the conscious components of recognition memory: Remembering is faster than knowing. *Consciousness & Cognition*, **15**, 147-162.
- 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.
- EFRON, B., & TIBSHIRANI, R. J. (1993). An introduction to the bootstrap. New York: Chapman & Hall.
- 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). Oxford: Oxford University Press.

- HEATHCOTE, A., BROWN, S., & COUSINEAU, D. (2004). QMPE: Estimating Lognormal, Wald, and Weibull RT distributions with a parameter-dependent lower bound. *Behavior Research Methods, Instruments, & Computers*, 36, 277-290.
- HEATHCOTE, A., POPIEL, S. J., & MEWHORT, D. J. K. (1991). Analysis of response time distributions: An example using the Stroop task. *Psychological Bulletin*, 109, 340-347.
- HOCKLEY, W. E. (1984). Analysis of response time distributions in the study of cognitive processes. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, 10, 598-615.
- HOHLE, R. H. (1965). Inferred components of reaction times as functions of foreperiod duration. *Journal of Experimental Psychology*, 69, 382-386.
- KUČERA, H., & FRANCIS, W. N. (1967). Computational analysis of presentday American English. Providence, RI: Brown University Press.
- MURDOCK, B. B., JR. (1985). An analysis of the strength-latency relationship. *Memory & Cognition*, 13, 511-521.
- MURDOCK, B. [B., Jr.] (2006). Decision-making models of remember– know judgments: Comment on Rotello, Macmillan, and Reeder (2004). Psychological Review, 113, 648-656.
- Petrusic, W. M., & Baranski, J. V. (2003). Judging confidence influences decision processing in comparative judgments. *Psychonomic Bulletin & Review*, **10**, 177-183.
- RAJARAM, S. (1996). Perceptual effects on remembering: Recollective processes in picture recognition memory. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, 22, 365-377.
- RATCLIFF, R. (1993). Methods for dealing with reaction time outliers. *Psychological Bulletin*, **114**, 510-532.
- RATCLIFF, R., & MURDOCK, B. B., Jr. (1976). Retrieval processes in recognition memory. *Psychological Review*, 83, 190-214.
- REDER, L. M., NHOUYVANISVONG, A., SCHUNN, C. D., AYERS, M. S., ANG-STADT, P., & HIRAKI, K. (2000). A mechanistic account of the mirror effect for word frequency: A computational model of remember–know judgments in a continuous recognition paradigm. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, 26, 294-320.
- ROTELLO, C. M., MACMILLAN, N. A., & REEDER, J. A. (2004). Sum—difference theory of remembering and knowing: A two-dimensional signal-detection model. *Psychological Review*, 111, 588-616.
- ROUDER, J. N., LU, J., SPECKMAN, P., SUN, D., & JIANG, Y. (2005). A hierarchical model for estimating response time distributions. *Psycho-nomic Bulletin & Review*, 12, 195-223.
- TULVING, E. (1985). Memory and consciousness. Canadian Psychology, 26, 1-12.
- VAN ZANDT, T., & MALDONADO-MOLINA, M. M. (2004). Response reversals in recognition memory. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, 30, 1147-1166.
- WIXTED, J. T. (2007). Dual-process theory and signal-detection theory of recognition memory. *Psychological Review*, 114, 152-176.
- 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. (2001). Consciousness, control, and confidence: The 3 Cs of recognition memory. *Journal of Experimental Psychology: General*, **130**, 361-379.

NOTES

- 1. Response reversals occurred on fewer than 2% of the trials, were distributed across conditions, and will not be discussed further. Those trials were excluded from all the reported analyses.
- 2. An eight-button response box was used in Experiment 2, with similar results.
- 3. d_a is a measure of sensitivity that takes account of the different variances of the target and lure distributions. When the strength distributions have equal variance, d_a equals d'.
- 4. Given the small sample sizes in many cells, only six tests were run. Even so, the power of each comparison is low.
- 5. The sample size for Subject A's remember responses in the long condition is below the recommended minimum for QMPE, which is 40. All of these data were also fit using PASTIS (Cousineau & Larochelle, 1997); the resulting parameter values were similar.

(Manuscript received April 27, 2006; revision accepted for publication January 16, 2008.)