

Relationship Between Measures of Working Memory Capacity and the Time Course of Short-Term Memory Retrieval and Interference Resolution

Ilke Öztekin
Brown University

Brian McElree
New York University

The response-signal speed–accuracy trade-off (SAT) procedure was used to investigate the relationship between measures of working memory capacity and the time course of short-term item recognition. High- and low-span participants studied sequentially presented 6-item lists, immediately followed by a recognition probe. Analyses of composite list and serial position SAT functions found no differences in retrieval speed between the 2 span groups. Overall accuracy was higher for high spans than low spans, with more pronounced differences for earlier serial positions. Analysis of false alarms to recent negatives (lures from the previous study list) revealed no differences in the timing or magnitude of early false alarms, thought to reflect familiarity-based judgments. However, analyses of false alarms later in retrieval indicated that recollective information accrues more slowly for low spans, which suggests that recollective information may also contribute less to judgments concerning studied items for low-span participants. These findings can provide an explanation for the greater susceptibility of low spans to interference.

Keywords: working memory capacity, short-term item recognition, interference, recent negative, speed–accuracy trade-off procedure

Measures of working memory capacity (WMC) have been found to predict performance in several cognitive tasks, including reading and language comprehension, vocabulary learning, writing, reasoning, problem solving, complex learning, and procedural skills (for a review, see Bors & MacLeod, 1996). Given that successful performance on many complex cognitive tasks requires maintenance and access to the products of prior perceptual and cognitive analyses, a reliable relationship between WMC measures and performance in these types of tasks is not surprising. However, it has proved somewhat more challenging to identify the particular memory operations that differences in WMC measures reflect. In the reported study, we investigated the relationship between WMC measures and retrieval operations—both automatic and controlled retrieval operations—that underlie recognition judgments of recent events.

Research investigating WMC effects has often used performance on WMC span tasks to form two contrast groups: High spans (HSs), those whose performance falls in the upper quartile, and low spans (LSs), those who are in the lower quartile (e.g., Conway et al., 2005; Unsworth, Heitz, Schrock, & Engle, 2005). A major finding to emerge from studies of memory performance is

that LSs appear to have a higher susceptibility to interference arising from distracting material (for a review, see Engle, 2002). For instance, Kane and Engle (2000) showed that LSs were more susceptible to proactive interference compared with HSs. Additionally, they found that when participants were required to perform a secondary task, performance of HSs decreased, whereas LSs' performance was unaffected. Accordingly, Kane and Engle suggested that HS use attentional control to compensate for the negative effects of proactive interference on their memory performance, whereas LSs do not normally allocate their attention to resist interference.

The importance of attentional control has been implicated in other tasks as well, such as the antisaccade task (Kane, Bleckley, Conway, & Engle, 2001), the Stroop task (Kane & Engle, 2003), and the dichotic listening task (Conway, Cowan, & Bunting, 2001). Collectively, these studies provide support for the controlled attention hypothesis (e.g., Engle, Kane, & Tuholski, 1999; Kane et al., 2001), which proposes that individual differences in WMC reflect limitations in attention allocation for the specific goals of the task at hand, especially in the face of interference or distraction. Specifically, controlled (or executive) attention refers to “a capability whereby memory representations are maintained in a highly active state in the presence of interference, and these representations may reflect action plans, goal states, or task-relevant stimuli in the environment” (Kane & Engle, 2002, p. 638). Accordingly, this account specifically predicts WMC effects to emerge in circumstances that require controlled processing.

One way to assess the respective contributions of automatic and controlled processes to a cognitive task is to measure performance as a function of processing time, as the output from automatic operations is typically available before the output of controlled operations (e.g., McElree, Dolan, & Jacoby, 1999; McElree & Doshier, 1989; Öztekin & McElree, 2007). However, to date, this

Ilke Öztekin, Department of Cognitive and Linguistic Sciences, Brown University; Brian McElree, Department of Psychology, New York University.

This research was supported by National Science Foundation Grant BCS-0236732 and National Institutes of Health Grant HD056200 to Brian McElree and by a graduate scholarship from the American Psychological Foundation and the Council of Graduate Departments of Psychology to Ilke Öztekin.

Correspondence concerning this article should be addressed to Ilke Öztekin, Department of Cognitive and Linguistic Sciences, Brown University, Box 1978, Providence, RI 02906. E-mail: ilke_oztekin@brown.edu

approach has not been applied to study the effects of individual differences in WMC. Here we report a time course investigation of how WMC impacts on the retrieval dynamics of short-term (probe) recognition (McElree & Doshier, 1989), using the response-signal speed-accuracy trade-off (SAT) procedure, which enables the independent and unbiased estimates of retrieval success and retrieval speed.

Measuring retrieval dynamics in the probe recognition task enables a direct assessment of whether differences in WMC are associated with differences in both retrieval speed and retrieval accuracy. To specifically assess how the contributions of automatic and controlled processes to memory performance may vary with WMC, we manipulated the recency of negative test probes (Monsell, 1978), a manipulation that has been widely used to investigate interference stemming from high episodic familiarity. Crucially for our purposes, manipulations of the recency of a negative probe served to put responses based on what is often regarded as an automatic assessment of familiarity in opposition to controlled retrieval operations (Jacoby, 1991; McElree et al., 1999). Specifically, a recently studied negative probe will have higher familiarity than a less recently studied (distant) negative probe, leading to an increased false-alarm (FA) rate. This higher FA rate can be corrected by the recovery of specific episodic (e.g., source) information about list membership.

A common finding in response time tasks (e.g., Monsell, 1978) is that correct rejections are longer and/or less accurate for recent negatives (RNs) than distant negatives (DNs). Importantly, in SAT time course studies (e.g., Hintzman & Curran, 1994; McElree & Doshier, 1989; Öztekin & McElree, 2007), RN probes have been found to induce high FA rates early in retrieval compared with DN probes. However, the elevated FA rates diminish later in retrieval when participants are able to recover more detailed episodic information (i.e., either that the probe was not a member of current study list or that it was studied on previous trial). This nonmonotonic FA function for RNs is consistent with two processes in opposition: Automatic assessments of familiarity engender high FA rates because of their high residual familiarity, which are then subsequently countered by controlled, strategic retrieval operations that serve to recover detailed episodic information.

The critical question we addressed in this study is whether WMC differentially impacts on early familiarity-based judgments or on the recovery of episodic retrieval processes that are operative later in retrieval. A fine-grain time course analysis has the potential to identify the underlying mechanisms responsible for the greater susceptibility of LS individuals to the presence of interference in a retrieval context. As noted, previous work suggests that WMC effects largely occur when controlled processing is required (e.g., Barrett, Tugade, & Engle, 2004; for reviews, see Engle, 2002; Unsworth & Engle, 2007). To the degree that familiarity judgments are predominantly based on automatic assessments of familiarity, and that judgments based on detailed episodic information require controlled recollective processes, we expected to see differences between HS and LS groups in later phases of retrieval when source information is used to modulate responses based on familiarity. Tracking differences between HS and LS groups across the time course of responses to recent and distant lures

enabled us to observe differences in both the magnitude and the timing of the two types of information accrual.

SAT Procedure

The SAT procedure can be used to measure the accuracy and speed of processing in a wide range of cognitive processes, including sentence comprehension (e.g., Foraker, & McElree, 2007; Martin & McElree, 2008; McElree, Foraker, & Dyer, 2003), visual attention (e.g., Carrasco, McElree, Denisova, & Giordano, 2003; McElree & Carrasco, 1999), and memory (reviewed in McElree, 2006). Application of SAT in the memory domain has largely focused on investigations of item recognition (e.g., Benjamin & Bjork, 2001; Hintzman & Curran, 1994; McElree & Doshier, 1989; Öztekin & McElree, 2007; Wickelgren, Corbett, & Doshier, 1980), although it has been implemented to characterize relational memory processes as well (e.g., temporal order, McElree & Doshier, 1993; spatial order, Grounlund, Edwards, & Ohrt, 1997; and *n*-back discriminations, McElree, 2001).

The main advantage of SAT is that it provides conjoint measures of the accuracy and the speed of processing. In contrast, response time measures from a reaction time task do not provide pure measures of processing speed. One problem with response time measures is that they are subject to SATs. More important, however, differences in the quality of memory representations can engender differences in response time, even when the underlying speed of information accrual is identical (e.g., Doshier, 1976, 1981; McElree & Doshier, 1989; Murdock, 1971; Ratcliff, 1978; Wickelgren, 1977; Wickelgren et al., 1980; for a review and discussion, see McElree, 2006). Response time measures are therefore less than optimal in applications such as the present when the research goal necessitates breaking apart processing speed from terminal accuracy and teasing apart the contributions of automatic and controlled processes—precisely what is required in the current investigation.

In the SAT procedure, participants are cued to respond to a response signal (a tone) presented at one of several (typically 6–7) times ranging from 40 to 3,000 ms after the onset of the probe. The time of the response signal is random on any trial, and participants are trained to respond within 100–300 ms of the tone. Varying the response signal across this range of times allowed us to measure the full time course of retrieval. This procedure enabled us to construct a retrieval function—accuracy as a function of processing time—for each condition of interest for each participant. SAT retrieval functions typically show an early period of chance performance, followed by a period of rapid increase in accuracy as retrieval time increases, and finally an asymptotic period of accuracy is reached, where additional retrieval time does not improve accuracy (illustrated in Figure 1A). The shape of the functions is usually well fit by an exponential approach to a limit. Three parameters describe these functions: (a) an asymptotic accuracy parameter revealing overall limitations of memory; (b) an intercept, indicating the point at which performance departs from chance; and (c) a rate of rise from chance to an asymptote. The asymptote parameter indicates the probability of successful retrieval, and the intercept and the rate parameters jointly constitute retrieval speed measures.

Method

Participants

Two-hundred and forty-three adults were screened in order to obtain WMC measures using the automated reading span and the automated operation span tasks (Conway et al., 2005; Unsworth, Heitz, Schrock, & Engle, 2005). Participants who scored in the upper quartile and lower quartile in both tasks constituted the HS and LS groups, respectively. LS and HS participants were then contacted for participation in the SAT experiment. Nine HS and 10 LS individuals agreed to participate in the experiment. For the screening session, 18 of the participants were paid for their time, and the remaining participants received credit for psychology classes via New York University's subject pool system. All par-

ticipants who took part in the experimental sessions were paid for their time.

Design and Stimuli

The experiment consisted of six 1-hr sessions, completed over a period of several weeks. Each session contained three 20-min blocks. Each block consisted of 168 experimental trials, in which participants studied a six-item list and were cued to respond to a recognition probe following a brief visual mask. Participants indicated whether the test probe was a member of the study list.

The design of the study was based on that of McElree (1998), which used categorized lists. Lists composed of instances from different categories enabled us to create negative probes that tested participants' ability to discriminate studied categories from unstudied categories and studied instances from semantically related but unstudied instances of a studied category. The stimulus set consisted of 36 categories, containing 21 words each, from the category norms of Van Overschelde, Rawson, and Dunlosky (2004). A study list for a trial was constructed by randomly selecting (without replacement) three members of a semantic category for the first three serial positions and randomly selecting (without replacement) three members of another category for the remaining three serial positions. Selection of categories excluded the ones that were used on the preceding trial. Positive probes were drawn equally often from one of the six serial positions of the current study list. One quarter of the negative probes consisted of new lures (NNs), lures that were not members of either of the two categories on the study list or words presented on the study list of the preceding trial. One quarter consisted of DNs from the first category (DN1): unstudied members of the same semantic category of the first three words in the study list. One quarter of the negative probes contained DNs from the second category (DN2): unstudied members from the same semantic category of the last three words in the study list. The remaining quarter of the negative

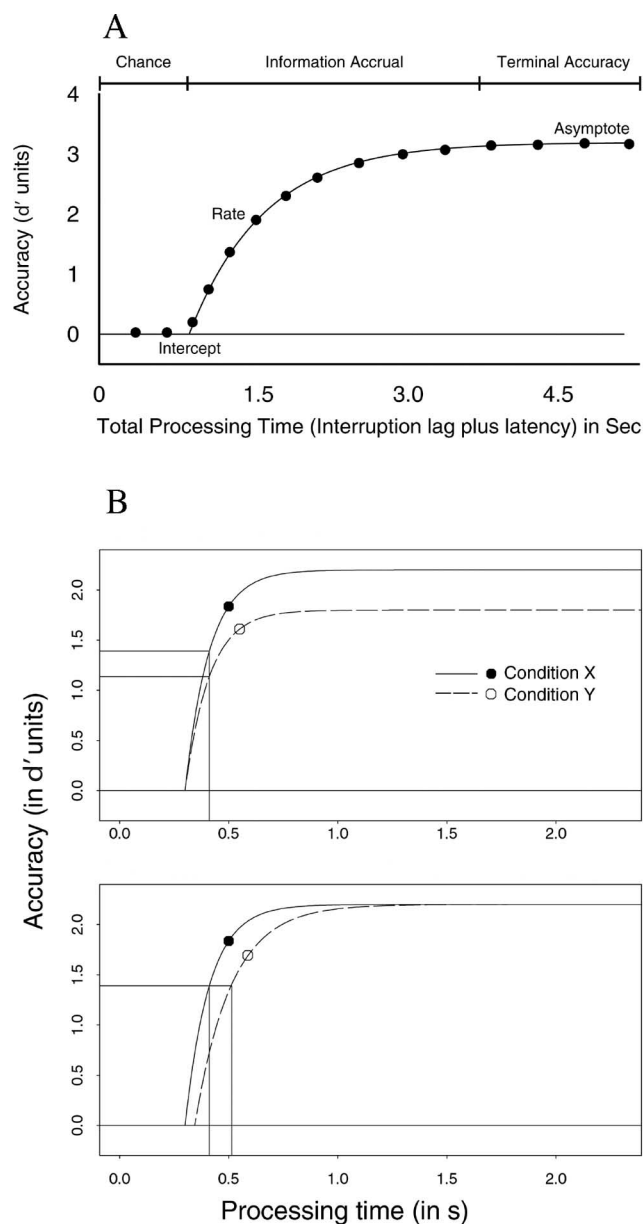


Figure 1. Illustration of hypothetical speed-accuracy trade-off (SAT) functions. (A) An example SAT function that shows how accuracy (in d' units) grows over processing time (in seconds). The SAT curve reflects three phases: a period in which performance is at chance (the departing point from chance is marked by the intercept parameter), followed by a period of information accrual (the rise of this information accumulation is reflected by the rate parameter of the SAT function), followed by the point at which the maximum level of accuracy is reached, where performance does not improve anymore (the asymptote parameter of the SAT function). (B) Examples of two SAT functions that illustrate differences in retrieval success versus retrieval speed. Both panels indicate accuracy measured (in d' units) plotted against total processing time (in seconds). The top panel shows a case in which two experimental conditions differ in probability of retrieval alone (e.g., a manipulation that affects memory strength alone). This difference is reflected in the asymptote parameter of the SAT function (i.e., Condition X has a higher asymptote than Condition Y), but both conditions have the same retrieval speed measures (i.e., the same intercept, the point at which performance departs from chance, and the same rate, which reflects the rate of information accrual). The bottom panel, on the other hand, illustrates a hypothetical case in which two experimental conditions differ in retrieval speed measures, displaying disproportional dynamics. That is, two functions reach a given proportion of their asymptote at different times.

probes consisted of RNs: lures drawn from studied words of the preceding trial. Note that as the selection of the categories of the current study list excluded those used on the preceding trial, RN probes were always from a category that had not been studied in the current trial. This ensured the isolation of the effects of residual familiarity from recent study, independent of semantic similarity effects.

This design structure resulted in 36 trials for each of the seven response deadlines for the six serial positions and 54 trials for each response deadline for the four lure types.

Procedure

F2 Figure 2 illustrates the sequence of events in a single trial: (a) A centered fixation point was presented for 500 ms. (b) Study words were then presented sequentially for 500 ms each. (c) The study list was followed by a visual mask, consisting of nonletter symbols for 500 ms. (d) Following the mask, the test word was presented for the duration of the response deadline. (e) At 43, 200, 300, 500, 800, 1,500, or 3,000 ms after the onset of the recognition probe, a 50-ms tone sounded to cue the participants to respond. (f) Participants indicated a yes–no recognition response as quickly as possible after the onset of the tone by pressing a key. (g) After indicating their response, participants were given feedback on their latency to respond. Participants were trained to respond within 300 ms of the tone.¹ They were informed that responses longer than 300 ms were too slow and responses under 100 ms were anticipations, and that both should be avoided. (h) After the latency feedback, participants were asked to give a confidence rating ranging from 1 (*low confidence*) to 3 (*high confidence*). The confidence ratings primarily served to enable participants to self-pace themselves through trials and were not analyzed. Participants initiated the next trial by pressing a key. Participants were allowed to take breaks between blocks.

Fn1

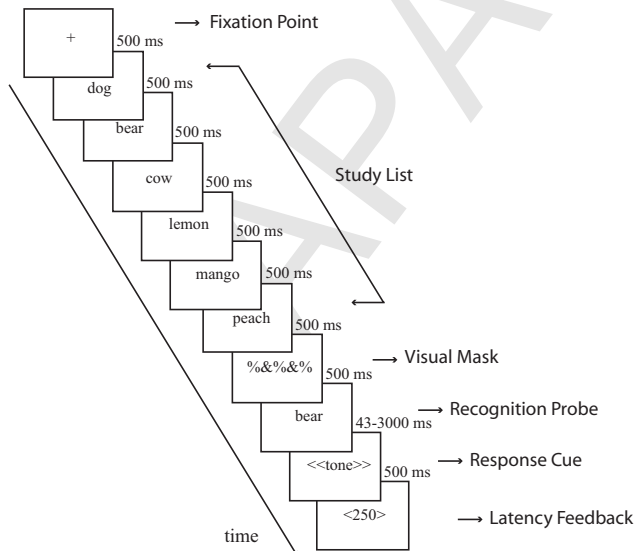


Figure 2. A sample trial sequence from the experimental procedure.

Results

For positive trials, each participant's hit rates were scaled against the FA rates to new lures (NNs) to obtain (equal-variance Gaussian) d' measures. To ensure that d' s were measurable, we adjusted perfect performance in any condition with a minimal corrections procedure, whereby hit rates higher than .99 were adjusted to .98 and FA rates lower than .01 were adjusted to .02, approximating the correction suggested by Snodgrass and Corwin (1988).

Asymptotic Accuracy

We averaged d' s for the last two response deadlines to obtain an empirical measure of asymptotic accuracy, which reflects the maximum amount of accuracy reached and is a measure of probability of successful retrieval (e.g., McElree, 2001; McElree & Doshier, 1989, 1993; Öztekin & McElree, 2007). Figure 3 illustrates the average asymptotic accuracy across the six serial positions for the HS and LS groups. A 2 (group [HS vs. LS]) \times 6 (serial position of test probe) mixed analysis of variance (ANOVA) conducted on the asymptotic d' s indicated a main effect of group: HS participants were more accurate than LS participants, $F(1, 17) = 8.332, p < .010$. There was also a reliable main effect of serial position, with more recent serial positions exhibiting higher accuracy, $F(1, 17) = 138.260, p < .001$. In addition, this analysis indicated a significant Group \times Serial Position interaction: LS participants were less accurate than HS participants for early positions, but this difference was less prominent as the test probe was more recent, $F(1, 17) = 7.718, p < .013$.

These asymptotic differences indicate that LS participants have lower overall probability of retrieval success in short-term recognition compared with HS individuals. Moreover, this difference in memory performance across the two groups is more prominent for the early members of the study list compared with more recent probes.

Retrieval Dynamics

We estimated the retrieval dynamics by fitting the individual participants' data and the average data (derived by averaging d' values for each condition across participants) with an exponential approach to a limit:

$$d'(t) = \lambda(1 - e^{-\beta(t - \delta)}), t > \delta, \text{ else } 0. \quad (1)$$

In Equation 1, $d'(t)$ is the predicted d' at time t ; λ is the asymptotic accuracy level reflecting the overall probability of recognition; δ is the intercept reflecting the discrete point when accuracy departs from chance ($(d') = 0$); and β is the rate parameter, which indexes the speed at which accuracy grows from chance to asymptote. Previous studies have indicated that this equation provides a good quantitative summary of the shape of the SAT functions (e.g., Doshier, 1981; McElree, 2001; McElree &

¹ Participants completed a 30-min practice session before the first experimental session to train for the SAT procedure. In addition, the trials in which participants took longer than 500 ms to respond, as well as trials in which participants responded before the tone, were excluded from analysis.

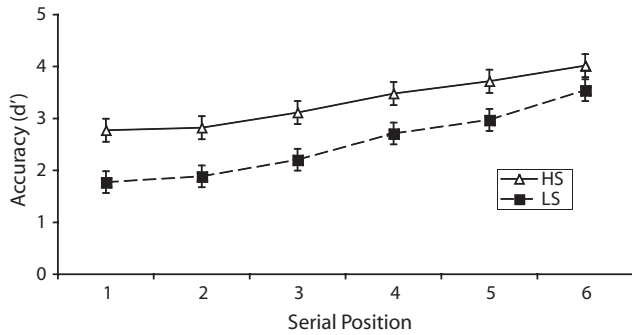


Figure 3. Empirical asymptotic accuracy (in d' units) measures, derived by averaging over the last two response deadlines, plotted across serial positions for the average high-span (HS) and low-span (LS) groups.

Dosher, 1989, 1993; Wickelgren & Corbett, 1977; Wickelgren et al., 1980).

The quality of the fits in the analyses reported below was examined by three criteria, which have been established in previous SAT studies: (a) the value of an adjusted R^2 statistic, which reflects the proportion of variance accounted for by a model, adjusted by the number of free parameters (Reed, 1973); (b) the consistency of the parameter estimates across participants; and (c) evaluation of whether the fit yielded systematic deviations that could be accounted for by additional parameters.

Retrieval Dynamics Within Lists

Initially, we evaluated whether SAT functions showed the same essential pattern that has been observed in other studies (e.g., McElree, 1996, 1998; McElree & Doshier, 1989, 1993; Öztekin & McElree, 2007; Wickelgren et al., 1980). To do so, we fit the SAT functions for the six serial positions with sets of nested models that systematically varied the three parameters of Equation 1. These models ranged from a null model in which all functions were fit with a single asymptote (λ), rate (β), and intercept (δ) to a fully saturated (18-parameter) model in which each function was fit with a unique asymptote, rate, and intercept.

Consistent with the results of prior studies, the data indicated retrieval dynamics differences due to fast-rising functions for Serial Position 6. The $6\lambda-2\beta-1\delta$ and $6\lambda-1\beta-2\delta$ models provided the best fit of the empirical data. The $6\lambda-2\beta-1\delta$ model allocated a separate asymptote to each serial position, one rate for Serial Positions 1–5, another rate for Serial Position 6 (the most recently studied item), and a common intercept for all the

six serial positions. This two-rate model significantly increased adjusted R^2 value compared with the $6\lambda-1\beta-1\delta$ model, $t(18) = 4.61, p < .05$. The increase in adjusted R^2 value resulted from a faster rate parameter for the last serial position. The rate in $1/\beta$ -ms units was 114 ms for Serial Position 6 versus 207 ms for other serial positions. The difference between the two rate parameters across participants was statistically significant, $t(18) = 5.66, p < .05$. Similarly, the $6\lambda-1\beta-2\delta$ model (that allocated a common rate to all serial positions, one intercept for Serial Positions 1–5, and another intercept for Serial Position 6) also reliably increased adjusted R^2 value compared with the $6\lambda-1\beta-1\delta$ model, $t(18) = 3.98, p < .05$. The intercept was 281 ms for Serial Position 6 versus 337 ms for other serial positions. The difference between the two intercept parameters across participants was statistically significant, $t(18) = -7.70, p < .05$.

For a subset of participants—five (out of nine) HS and two (out of 10) LS participants—there was evidence of a fast-rising function that included the last three serial positions on the list. For these participants, a model that allocated the faster rate to the last three items further increased the adjusted R^2 value from the $6\lambda-2\beta-1\delta$ and $6\lambda-1\beta-2\delta$ models reported above. As the most recent three positions shared the same semantic category, this finding is consistent with previous findings (e.g., McElree, 1998, 2006) that have demonstrated that the retrieval dynamics advantage for the last item can be increased to several items when the items can be grouped into a single chunk. It appears that only a subset of the participants strongly engaged in chunking by semantic category as a strategy for encoding the list.

Overall, the data replicate previous findings (McElree, 1996, 1998; McElree & Doshier, 1989; Öztekin & McElree, 2007; Wickelgren et al., 1980) regarding serial position effects on retrieval dynamics.

General Effects of WMC Measures on Accuracy and Retrieval Dynamics

We now turn to how WMC measures affect the retrieval dynamics of recognition memory over the short term. To conduct an initial comparison across the two WMC groups, and to determine whether the two groups differ in terms of retrieval dynamics, we averaged individual participants' and average (over participants) d' values across serial positions. These composite list SAT functions were fit with Equation 1 as described above. The resultant parameter estimates across participants are reported in Tables 1 and 2. Figure 4A illustrates the SAT functions for the average HS and LS

Table 1
Parameter Estimates From the Composite List Fit for High-Span Participants

Parameter	Average	Participant								
		1	2	3	4	5	6	7	8	9
λ	3.33	3.86	3.56	3.11	2.78	3.65	2.75	2.68	3.91	3.66
β	4.25	5.73	7.36	2.49	2.44	5.09	4.13	3.55	9.06	6.33
δ	0.309	0.352	0.345	0.294	0.256	0.321	0.333	0.374	0.271	0.352
R^2	.993	.983	.970	.894	.833	.959	.816	.953	.986	.995

Note. Average parameters are based on data averaged over participants (not the average of individual parameters).

T1-2
F4

Table 2
Parameter Estimates From the Composite List Fit for Low-Span Participants

Parameter	Average	Participant									
		1	2	3	4	5	6	7	8	9	10
λ	2.21	1.84	3.61	2.69	1.84	2.18	1.36	3.08	2.21	1.71	3.23
β	3.66	5.81	5.83	3.03	6.20	7.17	3.62	4.43	2.91	5.88	3.13
δ	0.281	0.280	0.295	0.306	0.288	0.281	0.394	0.308	0.299	0.329	0.253
R^2	.974	.833	.978	.973	.811	.912	.955	.916	.845	.611	.958

Note. Average parameters are based on data averaged over participants (not the average of individual parameters).

Fn2 data,² with smooth curves indicating the fitted exponential functions. As shown in the figure, LS participants had lower asymptotic accuracy than HS participants (2.21 vs. 3.33 for average LS and HS groups, respectively). Independent-samples *t*-test analysis conducted on the asymptote parameters across participants confirmed that this difference was reliable, $t(17) = 3.246, p < .005$.

Figure 4B presents the rate and intercept estimates across participants. There were no reliable differences in the rate or the intercept parameters across the two groups. Hence, the data failed

to reject the null hypothesis that HS and LS participants differ in retrieval speed, either in the rate of information accrual or when information first became available. As the speed differences occur because of differences operative early in retrieval, they may largely reflect the contributions of fast assessments of global memory strength or familiarity of the representation. Hence, the lack of a measurable difference in retrieval speed estimates across HS and LS participants is consistent with the controlled attention account, which asserts that differences across the two groups should occur only when controlled processing is required and not for decisions that are based on automatic processing, such as strength or familiarity assessments (e.g., Barrett, Tugade, & Engle, 2004).

Effects of WMC Measures on Serial Position

In addition to the list fits, we examined performance in more detail by deriving and modeling SAT functions for each of the six serial positions for individual participants' and average data for HS and LS groups. These functions were fit with Equation 1 as outlined above. Parameter estimates for each serial position across participants and the average data for the LS and HS groups are reported in Tables 3 and 4. Figure 5 shows the empirical SAT data and the fitted exponential functions for each of the six serial positions for the average HS and LS data. Between-groups comparisons conducted on the asymptote parameter of the serial position SAT functions indicated a lower asymptote for LS compared with HS participants for all serial positions: Special Position 1, $t(17) = 3.800, p < .001$; Special Position 2, $t(17) = 3.184, p < .005$; Special Position 3, $t(17) = 3.325, p < .004$; Special Position 4, $t(17) = 2.626, p < .018$; Special Position 5, $t(17) = 2.692, p < .015$; Special Position 6, $t(17) = 2.467, p < .025$. In addition, a 2 (group) \times 6 (serial position) mixed ANOVA was conducted on the asymptote estimates derived from the fits for each serial position function. Similar to the analysis reported on empirical *d'* values (see Asymptotic Accuracy section), this analysis indicated a reliable main effect of serial position, $F(1, 17) = 48.760, p < .001$, a reliable main effect of group, $F(1, 17) = 10.946, p < .004$, and a reliable Group \times Serial Position interaction, $F(1, 17) = 7.611, p < .013$. In line with the retrieval dynamics measures derived from the composite list SAT functions, there were no reliable differences

T3-4, F5

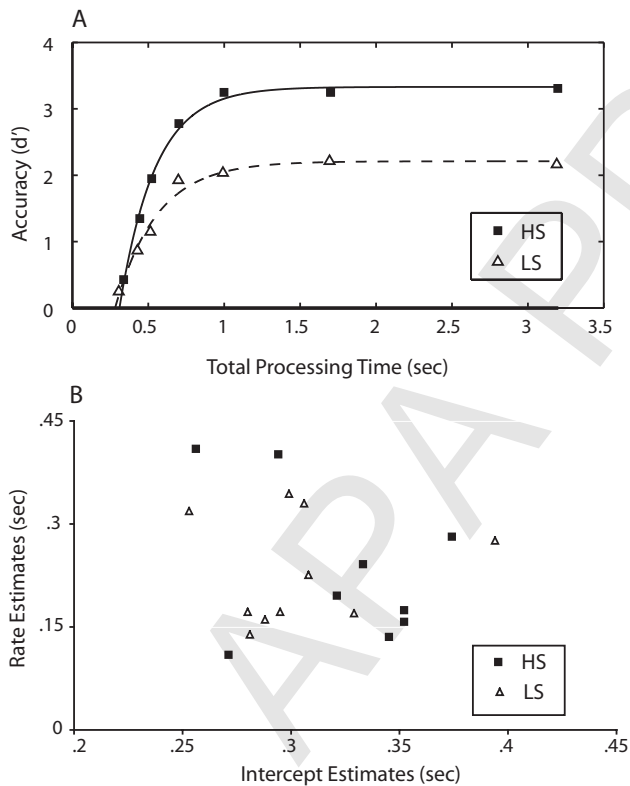


Figure 4. Composite list speed-accuracy trade-off (SAT) functions and model fits. (A) Accuracy (in *d'* units) for composite list (averaged over serial position of the test probe) SAT functions plotted against total processing time (in seconds) for the average high-span (HS) and low-span (LS) groups. The symbols indicate empirical data points, and the smooth curves indicate the model fits derived from Equation 1. (B) Retrieval dynamics estimates (the rate and intercept parameter estimates in seconds) across participants derived from the model fit presented in Figure 4A.

²Note that the illustrated SAT function is derived from fitting the average HS and LS data, rather than averaging over individual fits across participants.

Table 3
Parameter Estimates From the Serial Position (SP) Fits for High-Span Participants

Parameter	Average	Participant								
		1	2	3	4	5	6	7	8	9
SP 1										
λ	3.11	3.53	3.12	3.15	2.67	3.43	2.30	2.92	3.56	3.31
β	3.66	6.00	6.06	1.86	2.28	6.83	3.44	3.44	8.89	6.18
δ	0.331	0.381	0.369	0.323	0.270	0.418	0.391	0.415	0.284	0.387
R^2	.975	.935	.951	.858	.828	.935	.686	.920	.978	.985
SP 2										
λ	2.97	3.30	3.39	2.78	2.46	3.16	2.11	2.88	3.73	3.03
β	3.68	5.98	5.40	1.97	2.29	5.08	4.68	2.03	6.97	5.54
δ	0.320	0.401	0.324	0.343	0.217	0.352	0.408	0.297	0.285	0.360
R^2	.966	.988	.932	.747	.652	.939	.965	.913	.976	.959
SP 3										
λ	3.27	3.82	3.43	3.14	2.50	3.39	2.71	2.70	3.71	3.94
β	3.63	6.61	9.27	2.13	2.25	4.46	3.53	2.63	5.85	4.64
δ	0.322	0.377	0.368	0.335	0.277	0.340	0.357	0.369	0.266	0.360
R^2	.987	.983	.958	.816	.689	.904	.793	.889	.963	.961
SP 4										
λ	3.50	4.18	3.98	3.18	2.75	3.97	3.00	2.43	4.14	4.00
β	5.18	3.86	9.78	3.58	2.45	8.01	4.13	7.47	15.00	9.57
δ	0.316	0.287	0.355	0.324	0.286	0.340	0.332	0.428	0.274	0.377
R^2	.991	.938	.988	.891	.901	.952	.866	.938	.990	.982
SP 5										
λ	3.64	4.19	3.98	3.47	3.40	4.14	3.24	2.35	4.10	4.50
β	4.93	4.95	10.25	2.81	1.75	6.05	4.03	3.02	14.06	12.45
δ	0.303	0.289	0.363	0.304	0.200	0.317	0.273	0.276	0.277	0.401
R^2	.982	.941	.980	.855	.730	.999	.761	.925	.987	.932
SP 6										
λ	3.33	3.86	3.56	3.11	2.78	3.65	2.75	2.68	3.91	3.66
β	4.25	5.73	7.36	2.49	2.44	5.09	4.13	3.55	9.06	6.33
δ	0.309	0.352	0.345	0.294	0.256	0.321	0.333	0.374	0.271	0.352
R^2	.993	.983	.970	.894	.833	.959	.816	.953	.986	.995

Note. Average parameters are based on data averaged over participants (not the average of individual parameters).

across LS and HS groups in terms of the rate and the intercept parameters, suggesting that the two groups did not differ in retrieval speed.

FA Analyses

We next turn to our analyses of the negative probe conditions to examine whether and how HS and LS individuals differ in their response patterns to reject lures. Recall that four types of negative test probes were used: NNs from a semantic category other than the current study list, which had not been studied on the current or the preceding trial; DN1s from the first category (DN1), an unstudied item from the same semantic category of the first three serial positions of the study list; DN2s from the second category (DN2), an unstudied item from the same semantic category of the most recent three serial positions of the current memory set; and RNs, lures drawn from studied items in the preceding trial.

Descriptive analyses. Initial analyses were conducted on FA rates across the four lure types, the seven response deadlines, and the two groups. FA rates across conditions for HS and LS participants are illustrated in Figure 6.

A 2 (group) \times 7 (response deadline) \times 4 (type of lure) ANOVA was conducted on probability of falsely recognizing a lure as a member of the current study set. This analysis indicated a main effect of lure type, $F(3, 51) = 18.616, p < .001$. Additional

comparisons across the lure conditions revealed that participants were reliably more accurate in rejecting NN lures, compared with DN1 lures, $F(1, 17) = 9.444, p < .007$. DN1 probes were marginally more accurate than DN2 probes, $F(1, 17) = 3.763, p < .069$, and DN2 probes were more accurate than RN probes, $F(1, 17) = 4.941, p < .040$. Although the LS participants exhibited higher FA rates than HS participants in all four lure types, these differences did not reach statistical significance (mean FA rates for NN, DN1, DN2, and RN probes across the LS and HS groups were .163 and .117, respectively, for NN, .185 and .136 for DN1, .233 and .152 for DN2, and .256 and .200 for RN).

In addition, there was a reliable interaction between condition (RN vs. NN) and lag, $F(6, 102) = 28.254, p < .001$: Participants had a higher tendency to false-alarm to recent as compared with new lures early in retrieval, but the differences in FA rates diminished later in retrieval (see Figure 6). When the two false rates are directly scaled against each other, as in the next section, this interaction results in a nonmonotonic FA rate pattern for rejecting recently studied lures across retrieval time. This type of difference scaling directly shows the greater tendency to false-alarm to recently studied lures early in retrieval, which decreases with additional retrieval time. This pattern is consistent with that of previous research (e.g., Hintzman & Curran, 1994; McElree et al., 1999; McElree & Doshier, 1989; Öztekin & McElree, 2007) and impli-

Table 4
Parameter Estimates From the Serial Position (SP) Fits for Low-Span Participants

Parameter	Average	Participant									
		1	2	3	4	5	6	7	8	9	10
SP 1											
λ	1.95	1.36	3.34	2.35	1.57	1.83	0.698	2.89	1.59	1.25	2.61
β	3.49	2.09	4.56	2.74	4.06	4.49	15.00	3.72	15.00	15.00	3.03
δ	0.305	0.307	0.301	0.366	0.242	0.263	0.526	0.320	0.510	0.409	0.290
R^2	.959	.768	.940	.967	.568	.884	.895	.750	.832	.665	.910
SP 2											
λ	1.95	0.99	3.12	2.43	1.48	1.98	0.789	2.58	1.92	1.17	3.10
β	3.31	8.03	6.31	3.73	3.00	5.10	15.00	4.13	1.70	14.99	2.40
δ	0.296	0.331	0.358	0.406	0.277	0.269	0.526	0.322	0.227	0.505	0.299
R^2	.944	.244	.975	.895	.689	.861	.791	.867	.639	.501	.842
SP 3											
λ	2.17	1.32	3.55	2.29	1.72	1.98	1.08	2.99	1.89	1.69	3.28
β	3.70	3.35	4.26	2.84	8.00	5.74	2.76	5.03	2.40	3.09	3.20
δ	0.299	0.222	0.267	0.307	0.298	0.285	0.395	0.358	0.303	0.317	0.299
R^2	.958	.410	.937	.942	.563	.851	.751	.969	.834	.783	.986
SP 4											
λ	2.65	1.99	4.02	2.83	1.98	2.24	2.75	3.20	2.44	2.20	3.39
β	4.21	11.09	7.19	4.08	6.67	6.79	1.10	3.71	2.95	6.52	3.92
δ	0.294	0.376	0.306	0.409	0.305	0.287	0.267	0.285	0.276	0.326	0.260
R^2	.984	.901	.933	.956	.930	.793	.959	.810	.851	.394	.915
SP 5											
λ	2.86	2.51	3.93	2.97	2.84	2.46	1.86	3.46	2.58	2.23	3.87
β	4.71	10.06	9.71	4.65	2.61	11.23	3.67	4.09	8.56	9.14	3.56
δ	0.295	0.332	0.301	0.359	0.275	0.280	0.401	0.264	0.430	0.345	0.268
R^2	.994	.910	.927	.975	.823	.919	.890	.715	.886	.669	.962
SP 6											
λ	3.33	3.57	4.14	3.91	2.50	3.09	2.69	4.11	3.22	2.46	3.98
β	4.99	6.59	9.89	3.06	10.88	11.24	4.35	5.47	3.66	6.34	3.75
δ	0.267	0.247	0.280	0.232	0.272	0.271	0.386	0.299	0.288	0.321	0.240
R^2	.995	.874	.996	.852	.673	.858	.839	.957	.954	.602	.971

Note. Average parameters are based on data averaged over participants (not the average of individual parameters).

icates the contribution of two types of information accrual to recognition memory judgments: global strength (or familiarity) information, which is dominant early in retrieval leading to high FA rates, and recollective information, which becomes available later in retrieval via controlled, strategic episodic retrieval attempts. Accrual of the latter can diminish the FA rates if relevant episodic information is recovered (i.e., source information in this study). Below, we further investigate this interaction across our LS and HS participants.

FA scaling. To examine how WMC measures impact on the global assessments of strength or familiarity of an item (often regarded as an automatic process) and the recovery of detailed episodic information (often thought to be driven by controlled recollective processes), we computed FA difference scores between our RN probes and new lures—FA(RN)—FA(NN)—at each of the seven response deadlines. This measure allowed an unbiased measure of performance by factoring out participants' bias to judge an item as a member of the study list (e.g., tendency to respond yes more often than no, regardless of the type of test probe). Hence, the obtained FA difference scores served to reflect a pure measure of tendency to false-alarm because of high residual familiarity of the RN probes. Crucially, this measure enabled us to examine the differential impact of WMC measures on judgments based on familiarity and the accrual of detailed episodic (recollec-

tive) information, independent of a possible general tendency for LS participants to false-alarm more than HS participants.

Figure 7 illustrates the FA difference scores for the average LS and HS data. Note that with this scaling, higher scores indicate a higher tendency to false-alarm to the RN probes. The figure indicates that for both the HS and LS groups, the FA difference scores increase early in retrieval and then diminish later in retrieval. This nonmonotonic pattern indicates that the information basis for the recognition memory judgments has shifted across retrieval and is consistent with predictions of dual-process theories of recognition memory: The early high FA rates indicate the contribution of familiarity (as the RN probe has been studied on previous trial, it has high residual familiarity compared with the new lure). The observed reduction in FA rates later in retrieval suggests accrual of new information that contributes to the recognition judgments, presumably reflecting source or list-specific information (i.e., the fact that the RN probe was studied on previous trial, or that it was not a member of the current study list)—recovered by what is viewed as a recollective process in dual-process models of recognition memory (for a review, see Yonelinas, 2002). This biphasic nature of FA rates to RN probes reported in Figure 7 is consistent with that of previous SAT studies that have put familiarity and recollective information in opposition to each other

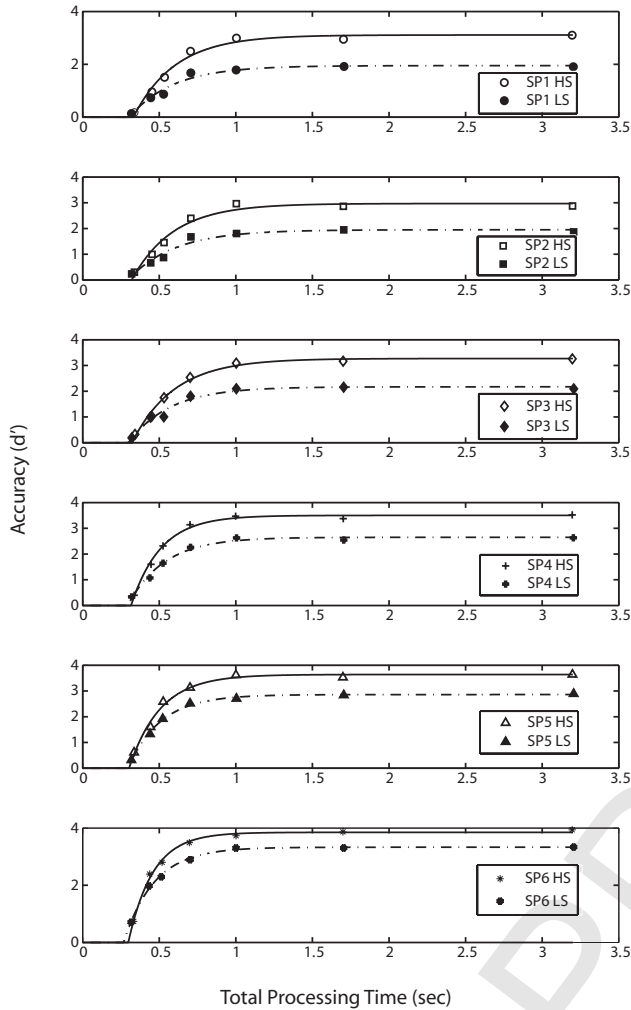


Figure 5. Accuracy (in d' units) for each serial position (SP) in the study list, plotted against total processing time (in seconds) for the average high-span (HS) and low-span (LS) groups. The symbols indicate empirical data points, and the smooth curves indicate the model fits derived from Equation 1.

(e.g., McElree et al., 1999; McElree & Doshier, 1989; Öztekin & McElree, 2007; Zeelenberg, Wagenmakers, & Shiffrin, 2004; for a complete review, see McElree et al., 1999).

Of particular interest here is the differences in magnitude or timing of the early familiarity-based judgments, as well as in the timing and accrual of recollective, episodic information across our HS and LS groups. Figure 7 shows that the diminish in FA rate occurs considerably later for the LS group compared with the HS group, suggesting that the accumulation of detailed episodic information—presumably recovered by controlled retrieval operations—might have started later and/or completed more slowly than for the HS group. However, the figure further indicates that given enough time (about 3.2 s), both groups reach comparable accuracy, suggesting that the two groups do not differ in the maximum level of accuracy reached after completion of the controlled retrieval process that accesses relevant episodic information.

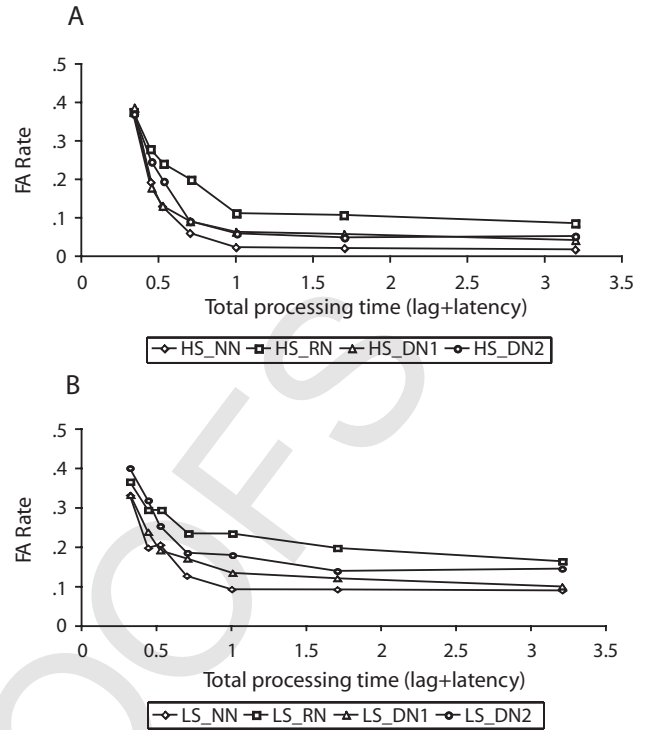


Figure 6. False-alarm (FA) rates plotted against total processing time (in seconds) for the lures. (A) FA rates for average high-span (HS) group. (B) FA rates for average low-span (LS) group. NN = new negative; RN = recent negative; DN1 = distant negative from the first category; DN2 = distant negative from the second category.

To quantify this pattern, and to test the observed differences across the two groups in Figure 7, we fit the FA difference scores with a model that explicitly assumes that retrieval shifts from one source of information to another source across processing time. This type of model is formally equivalent to a two-process retrieval model. Such a two-process retrieval model for SAT was originally proposed by Ratcliff (1980) and was later adapted to the exponential form (McElree & Doshier, 1989), which we apply to our current data set:

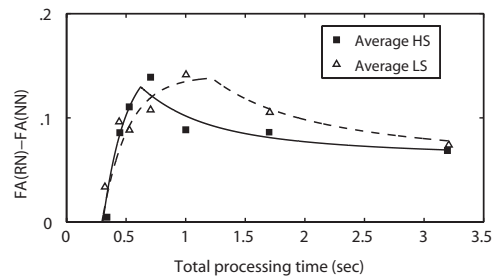


Figure 7. Difference in false-alarm (FA) rates across the recent negative (RN) and new negative (NN) probes plotted against total processing time (in seconds) for average high-span (HS) and low-span (LS) groups. The symbols indicate empirical data points, and the smooth curves indicate the model fit derived from Equation 2.

$$\begin{aligned}
 & \lambda(1 - e^{-\beta(t - \delta_1)}), \text{ for } \delta_1 < t < \delta_2 \\
 \text{FA}_{\text{diff}}(t) = & \lambda_2 + (\lambda_1 - \lambda_2)(\delta_2 - \delta_1)/(t - \delta_1) \\
 & \times (1 - e^{-\beta(t - \delta_1)}), \text{ for } t \geq \delta_2. \quad (2)
 \end{aligned}$$

Equation 2 states that during the initial retrieval period ($\delta_1 < t < \delta_2$), accuracy depends on accrual of one type of information, presumably familiarity information. During this initial period, accuracy is modeled by the top portion of Equation 2, a simple exponential approach to an asymptote (λ_1). At time δ_2 , a second source of information starts to contribute to the recognition memory judgments. This source of information could arise from the output from a second process (e.g., a recollective operation that accesses detailed episodic information). The accrual of this second type of information leads to the change in retrieval, shifting the asymptote from λ_1 to λ_2 . The bottom portion of Equation 2 states that response accuracy gradually shifts to the new asymptote (λ_2) starting at time δ_2 .

To test whether WMC measures affect familiarity-based responses or the recovery of more detailed episodic information, or both, we fit each participant's and the average LS and HS data with Equation 2 and compared the asymptote (λ_1, λ_2) and intercept (δ_1, δ_2) parameters that correspond to each. Specifically, if the two groups differ in the timing or the magnitude of responses based on familiarity information, we should see differences in, respectively, the familiarity asymptote (λ_1) and/or the familiarity intercepts (δ_1) across HS and LS participants. If, on the other hand, they differ in the timing or the magnitude of responses based on controlled processes that serve to recover detailed episodic (source) information, then we should see differences in the intercept parameter δ_2 , which indicates the point at which relevant episodic information first becomes available, and/or the asymptote parameter λ_2 , which reflects the maximum level of accuracy reached upon accrual of relevant episodic (source) information. Tables 5 and 6 present the parameter estimates from the fits of the average HS and LS data, and the individual participants' data in each group. The smooth functions in Figure 7 indicate the fits to the average HS and LS data.

Before examining differences across the two groups, we initially conducted comparisons between the λ_1 and λ_2 estimates and between the δ_1 and δ_2 estimates across all our participants (collapsing over group), which examines whether the observed data pattern is consistent with the model depicted in Equation 2. Paired t tests across these parameter estimates indicated that λ_1 , reflecting

the familiarity asymptote, was reliably higher than λ_2 , the asymptote reflecting the level of accuracy reached after recovery of detailed episodic information, $t(18) = 3.624, p < .002$, with 16 out of 19 participants showing this ordering. The higher λ_1 indicates the nonmonotonic nature of the functions, and it suggests an early intrusion of familiarity information, which is corrected later in retrieval, presumably with the accrual of specific episodic information. Additionally, δ_1 , reflecting the familiarity intercept (i.e., the point when familiarity information first becomes available), occurred earlier in retrieval than δ_2 , reflecting the intercept of the accrual of specific episodic information, $t(18) = -4.565, p < .001$, with 17 out of 19 participants showing this ordering. Hence, the data and the parameter estimates are consistent with the contribution of two types of information accrual to the recognition memory judgments, as represented in Equation 2.

Given the applicability of a two-process model, we next examined differences in the estimated parameters across HS and LS participants. These analyses enabled us to investigate whether WMC measures are related either to the magnitude and/or timing of familiarity-based early responses or to the magnitude and/or timing of episodic information.

Consistent with the average data and model fits illustrated in Figure 7, an independent-samples t -test comparison across the LS and HS participants indicated that neither the familiarity asymptote (λ_1) nor the familiarity intercept (δ_1) estimates differ across the two groups ($p > .1$), suggesting that HS and LS groups did not differ in either point when they began to false-alarm more to RN than to NN probes or the degree to which they false-alarmed to the two probe types.

Although the two groups do not appear to differ in their familiarity intercept estimates, inspection of the fits of the average data in Figure 7 might suggest that the rate at which familiarity information accrues is slower for the LS than the HS group. Equation 2 makes the simplifying assumption that the rate at which familiarity information accrues is the same as the rate at which episodic information accrues (both estimated by β). This assumption facilitates the recovery of stable parameter estimates by limiting the tendencies for parameter trade-offs that typically result from dual-process models assuming separate intercepts, rates, and asymptotes for the two forms of information (McElree & Doshier, 1989). Although the rate estimates from the above fit did not differ across the two groups, $t(17) = 0.18, p > .86$, fitting the data with a common rate could conceal differences between the HS and LS groups. To test whether there might be differences

Table 5
Parameter Estimates for the Dual-Process Fit for High-Span Participants

Parameter	Average	Participant								
		1	2	3	4	5	6	7	8	9
λ_1	0.160	0.310	0.190	0.290	0.248	-0.150	0.078	0.200	0.070	0.154
λ_2	0.058	0.176	0.030	0.004	-0.006	0.140	0.170	0.030	0.055	-0.002
β	5.25	5.51	7.49	3.97	3.13	7.56	2.92	4.78	10.00	8.70
δ_1	0.306	0.292	0.292	0.371	0.258	0.319	0.198	0.370	0.266	0.351
δ_2	0.623	0.674	0.514	0.434	0.449	0.300	1.06	0.639	0.300	0.655
R^2	.998	.951	.994	.963	.932	.975	.928	.985	.987	.995

Note. Average parameters are based on data averaged over participants (not the average of individual parameters).

Table 6
Parameter Estimates for the Dual-Process Fit for Low-Span Participants

Parameter	Average	Participant									
		1	2	3	4	5	6	7	8	9	10
λ_1	0.140	0.120	0.240	0.230	0.290	0.130	0.163	0.600	0.130	0.170	0.060
λ_2	0.050	-0.021	0.055	0.040	0.030	-0.065	-0.015	0.125	-0.070	0.000	0.190
β	4.67	7.66	8.76	3.90	6.72	7.73	3.91	5.73	8.70	5.13	3.72
δ_1	0.297	0.277	0.296	0.321	0.292	0.293	0.393	0.323	0.409	0.347	0.289
δ_2	1.22	1.72	0.493	0.389	0.623	1.36	1.47	0.366	1.60	1.31	0.783
R^2	.994	.957	.959	.988	.938	.989	.932	.936	.954	.827	.973

Note. Average parameters are based on data averaged over participants (not the average of individual parameters).

in rates at which the two types of information accrue across the LS and HS groups, we fit a variant of the model in Equation 2 that allowed for two rates (β s), one for familiarity information and for the late accruing episodic information. The quality of the fit for this model was substantially less than the simpler single-rate model in Equation 2. Specifically, for the average LS data, R^2 dropped from .99 to .91 in fits of the average LS data and from .99 to .86 in fits of the average HS data, and this fit yielded unstable parameter estimates.

Hence, WMC measures did not have a measurable relationship to early familiarity-based responses. The λ_2 asymptote parameter, reflecting the maximum level of accuracy reached after the accumulation of specific episodic information, also did not differ across the two groups. However, there was a reliable difference between the two groups in the timing of when detailed episodic information began to accrue. The δ_2 intercept parameter, reflecting the onset of when the corrective influence of specific episodic information became evident, was significantly slower for LS participants than for HS participants, $t(17) = -2.352, p < .031, d = 1.10$. In addition, a default Bayesian t test (Rouder, Speckman, Sun, Morey, & Iverson, 2009) using the “unit information prior” yielded a Bayes factor of 0.3, indicating that the data are approximately 3 times more likely to have come from the alternative than the null hypothesis. In the average data, δ_2 estimates were 623 ms for the HS group compared with 1,220 ms for the LS group.

A single-process model could account for the nonmonotonic pattern of FA rates by assuming that RNs are rejected more slowly and thus result in a longer time for reducing the FA rates for RNs (for such an account, see, e.g., Brockhoff & Lamberts, 2000). However, such a model is less able to account for related nonmonotonic patterns with crossover effects engendered by manipulations of a single variable. For example, McElree et al. (1999) found that repeated study of a lure produces higher FA rates early in retrieval than a once-studied lure but lower FA rates later in retrieval. Crucially, here a single-process model cannot explain the fact that WMC is related to the speed of rejecting lures but not related to the speed with which studied items are retrieved. Hence, these findings suggest that WMC selectively impacts on the accrual of detailed episodic information later in retrieval, which maybe recovered by controlled recollective processes. We found no evidence to indicate that WMC differences are related to the processes responsible for early strength- or familiarity-based responses.

Discussion

General Effects of WMC Measures on Retrieval Success and Retrieval Speed

In both the analyses of the composite list SAT functions (viz., averaged over serial position) and the individual serial position SAT functions, we found that LS participants had lower accuracy than HS participants. Reliably lower performance was evident in both the empirical measures of asymptotic accuracy (average of the two longest response deadlines) and the associated asymptotic estimates derived from fits of the exponential retrieval equation (Equation 1). These differences in retrieval accuracy are not altogether surprising, as participants were partitioned into HS and LS groups in part by the accuracy of their retrieval, albeit in recall rather than recognition tasks. Perhaps more notable is the absence of any concomitant differences in retrieval speed: Retrieval dynamics measures—the intercept parameter that indicates the point at which information first becomes available and the rate parameter that reflects the rate of information accrual—did not differ across the two groups. Prima facie, the absence of dynamics differences suggests that WMC measures are not related to the speed with which participants can gain access to a representation in memory. However, this claim must be qualified by the analyses of the rejection of negative probes, discussed more fully below, which indicates that LS participants recovered detailed episodic information more slowly than HS subjects.

It is often assumed that recognition judgments reflect the contributions of two sources of information: an assessment of the overall quality of the match of a test probe to representations in memory (often viewed as an assessment of familiarity) and the recovery of detailed episodic or contextual information (such as source information). The latter has commonly been viewed as a recollective process in dual-process theories of recognition (for a review, see Yonelinas, 2002). Time course investigations of memory retrieval have consistently found that familiarity information is available earlier than detailed episodic information (e.g., Hintzman, Caulton, & Levitin, 1998; Hintzman & Curran, 1994; McElree et al., 1999; McElree & Doshier, 1989; Öztekin & McElree, 2007). We suggest that the absence of WMC-related differences in retrieval speed for cases when the probe matches an item from the study list indicates that HS and LS participants are equally facile at recovering global assessments of the familiarity of the test probe. As discussed in the next section, this claim is

supported by analyses of the FA rates for negative probes, which likewise indicate that HS and LS participants do not differ in judgments based on familiarity. Crucially, it is these fast assessments of familiarity that primarily determine performances in the early (preasymptotic) portions of the SAT functions and hence control the dynamics—the intercept and rate estimates—of the functions (see McElree et al., 1999). To the degree that the recovery of familiarity is a largely automatic process, the absence of an effect of WMC on familiarity-based judgments is consistent with the controlled attention view (e.g., Engle, 2002), which asserts that WMC effects occur during circumstances that require controlled processing. It is also consistent with previous research (e.g., Oberauer, 2005) suggesting that WMC is related to the efficiency of recollection but not familiarity.

Differential Impact of WMC Measures on Automatic and Controlled Processes That Contribute to Short-Term Recognition

Manipulating the recency of negative probes enabled us to isolate and track the time course of responses based on familiarity and those based on the recovery of detailed episodic information. Consistent with our interpretation of the data from positive probes, the comparison of FA rates to RNs and DNs across the time course of recognition indicated that HS and LS groups did not differ in the magnitude or timing of the early familiarity-based judgments. That is, when RNs were scaled against DNs, thereby isolating differences in FAs between the two negative probe types, HS and LS participants did not differ in the point when the increased FA rate for RNs first appeared (δ_1 in Equation 2) or in the maximal FA rate observed (λ_1 in Equation 2).

Notably, however, we found that the point when participants began to correct the high FA rate for RNs was substantially earlier for the HS group than the LS group. The δ_2 parameter in dual-process model (Equation 2) estimates the point when detailed episodic information is first used to correct responses based on familiarity. Our data indicate that this inflection point in the FA functions was on average 600 ms later for LS than HS participants. Although neither group completely overcame the misleading effects of recent study, LS participants were eventually able to reduce their FA rate to the level of HS participants.

The delayed inflection point for LS participants unequivocally indicates that those with low WMC are slower at using episodic information to correct for the misleading effects of familiarity. There are at least three potential reasons for why this might be the case: (a) LS participants might retrieve episodic information at slower rates than HS participants; (b) LS participants might retrieve episodic information at rates comparable to those of HS participants, but they take longer to initiate controlled retrieval operations; or (c) LS participants may initiate retrieval operations and recover episodic information at times comparable to those of HS participants, but they may take longer to resolve the conflict between that information and the high familiarity of an RN. Any of these three reasons in isolation or in combination is fully compatible with our findings, but we believe that Explanations b and c are the most plausible when viewed in a wider context.

Although it is possible that the intrinsic rate at which LS participants retrieve episodic information is substantially slower than the rate for HS participants, the magnitude of the difference

between HS and LS—on average 600 ms—far exceeds the differences that are typically observed from various experimental manipulations of factors that affect retrieval³ or the differences that are observed between individuals in various memory tasks. Additionally, even if global assessments of familiarity and detailed episodic information are recovered with qualitatively different operations, both require access to a memory representation, and without a detailed characterization of the operations used to recover the two forms of information, it is unclear why the former but not the latter should vary with WMC. We believe that Explanations b and c are more consistent with the evidence from a range of tasks (reviewed in Engle, 2002) suggesting that WMC effects are observed when executive or control processes are essential for performing at high-accuracy levels. Finally, the view that WMC may reflect the ability to maintain context bindings (Oberauer, 2005) is also consistent with our findings, as recollective information may often rest on content–context bindings, whereas familiarity is commonly assumed to reflect the strength of an individual item independent of its context.

Explanations b and c do not assume that WMC predicts the speed of episodic retrieval per se but rather that it is related to the speed with which participants initiate episodic retrieval operations to counter the misleading effects of familiarity (b) or the speed with which they resolve the conflict between familiarity and episodic information (c). Our data do not discriminate between b and c, and so we refer to them jointly as a controlled, strategic retrieval operation. Crucially, however, both accounts appear consistent with the controlled attention hypothesis (e.g., Engle et al., 1999; Kane et al., 2001) that argues that low WMC reflects a limitation in attention allocation for the specific task goals, especially in the face of interference. These accounts attribute the WMC effects to controlled, strategic operations that are similar to those that may be involved in other tasks in which WMC effects have been observed, such as the antisaccade task (Kane et al., 2001), the Stroop task (Kane & Engle, 2003), and the dichotic listening task (Conway et al., 2001).

We also note that Unsworth and Engle (2007) have suggested that the operation span task (one of the two tasks used in this study to determine WMC) may reflect two components, namely, primary and secondary memory. Within such a framework, the differential effects of WMC on the accrual of recollective information would most plausibly have their locus in the secondary memory component, as this component is hypothesized to reflect controlled memory search (Unsworth & Engle, 2007).

Finally, we note that beyond providing a specific explanation for the greater susceptibility of LS participants to misleading familiarity information—namely, inefficient deployment of controlled retrieval processes to successfully resolve interference—the controlled, strategic retrieval hypothesis also provides an explanation of two other properties of our data. First, it provides an explanation

³ An exception might be the 500-ms differences in SAT intercept observed in McElree and Doshier's (1993) investigation of judgments of recency. However, those differences arose from a manipulation that varied the number of serial retrieval operations required to reach a judgment, from one to five hypothesized operations. As such, they are more properly viewed as arising from repetitive operations rather than intrinsic differences in one type of operation.

of the observed reduction in accuracy for LS participants for positive trials (discussed above). For the positive test probes, the recognition judgments can be based on both familiarity and episodic information. Although the contribution of familiarity and recollective information to studied probes cannot be isolated (as they both contribute to tendency for a positive response), if LS participants rely less on episodic information (e.g., as a result of recollective information accruing slower), then naturally their asymptotic levels of performance on positive trials would be less than those of HS participants, who can more effectively make use of episodic recollective information in their memory judgments. Additionally, this account can explain the observed Group \times Serial Position interaction. Our data indicated that the difference in accuracy across the two groups was more prominent for early serial positions in the study list. Recently studied items have higher familiarity than less recently studied items, so greater reliance on familiarity information will have less of an adverse effect. Additionally, episodic information might be more likely to be spontaneously recovered, without the need to engage in controlled, strategic retrieval operations, for recently studied items. In line with both claims, Öztekin and McElree (2007) found that the most recent three positions in six-item lists comparable to those of the current study were immune to proactive interference effects on asymptotic accuracy.

Controlled, Strategic Retrieval: Implications for Neural Accounts of WMC

The neural mechanisms that mediate interference resolution in the RN probe paradigm have been widely studied. Neuroimaging studies (e.g., Badre & Wagner, 2005; Jonides, Badre, Curtis, Thompson-Schill, & Smith, 2002; Jonides et al., 2000; Jonides, Smith, Marshuetz, Koeppe, & Reuter-Lorenz, 1998; Nelson, Reuter-Lorenz, Sylvester, Jonides, & Smith, 2003; Öztekin, Curtis, & McElree, 2009; for a review, see Jonides & Nee, 2006) have identified enhanced activation in left ventrolateral prefrontal cortex, namely the left inferior frontal gyrus (LIFG), for RN probes compared with unstudied probes in item recognition. It has also been shown that this effect is specific to the retrieval stage of the RN probe (D'Esposito, Postle, Jonides, & Smith, 1999). Additionally, patient work (e.g., Thompson-Schill et al., 2002) and repetitive transcranial magnetic stimulation investigations (e.g., Feredoes, Tononi, & Postle, 2006) have provided converging evidence for a direct role of LIFG in successful interference resolution in this paradigm.

Behavioral work suggests that resolving interference in the RN probes can be achieved via controlled retrieval processes that recover specific episodic information (such as source or list-specific information; e.g., McElree & Doshier, 1989; Öztekin & McElree, 2007). In addition, LIFG has been implicated in strategic retrieval of episodic information in the absence of interference (e.g., Dobbins, Rice, Wagner, & Schacter, 2003), and patients with left prefrontal cortex lesions show deficits in other tasks that require the recovery of detailed episodic information, such as source memory (e.g., Duarte, Ranganath, & Knight, 2005). Furthermore, neural activation in this region was found to be modulated by the amount of successive retrieval operations carried out to recover temporal order information in a short-term judgment-of-recency task (Öztekin, McElree, Staresina, & Davachi, 2008).

Accordingly, it has been suggested that the role of LIFG in resolving interference in the RN probe paradigm may be supporting the controlled, strategic retrieval operations that access relevant episodic information that can successfully resolve interference (e.g., Öztekin et al., 2009).

Our data indicate that LS and HS participants differed in the ability to deploy and use controlled, episodic retrieval operations necessary to resolve interference, but there was no evidence suggesting a measurable effect of WMC on the timing and magnitude of fast familiarity-based judgments. Hence, the current results provide converging evidence to studies that have implicated the importance of executive or controlled attention in yielding WMC effects and the greater susceptibility of LS individuals to interference (for an overview, see Engle, 2002), and provide further support for the contention that these effects may be mediated by the left ventrolateral prefrontal cortex (for similar arguments regarding the role of the prefrontal cortex in modulating WMC effects, see also Kane & Engle, 2002; Rosen & Engle, 1997). Future work investigating the role of this region in mediating WMC effects would be beneficial in advancing our understanding of the underlying mechanisms that lead to individual differences in WMC. In addition, the time course pattern implicated in our study can be followed up by methods that have better temporal resolution than functional magnetic resonance imaging, such as electroencephalography and magnetoencephalography.

References

- Badre, D., & Wagner, A. D. (2005). Frontal lobe mechanisms that resolve proactive interference. *Cerebral Cortex*, *15*, 2003–2012.
- Barrett, L. F., Tugade, M. M., & Engle, R. W. (2004). Individual differences in working memory capacity and dual-process theories of the mind. *Psychological Bulletin*, *130*, 553–573.
- Benjamin, A. S., & Bjork, R. A. (2001). On the relationship between recognition speed and accuracy of words rehearsed via rote versus elaborative rehearsal. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *26*, 638–648.
- Bors, D. A., & MacLeod, C. M. (1996). Individual differences in memory. In E. L. Bjork & R. A. Bjork (Eds.), *Handbook of Perception and Cognition: Vol. 10. Memory* (pp. 411–441). San Diego, CA: Academic Press.
- Brockhoff, N., & Lamberts, K. (2000). A feature-sampling account of the time course of old–new recognition judgments. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *26*, 77–102.
- Carrasco, M., McElree, B., Denisova, K., & Giordano, A. M. (2003). The speed of visual information processing increases with eccentricity. *Nature Neuroscience*, *6*, 669–670.
- Conway, A. R. A., Cowan, N., & Bunting, M. F. (2001). The cocktail party phenomenon revisited: The importance of working memory capacity. *Psychonomic Bulletin & Review*, *8*, 331–335.
- Conway, A. R. A., Kane, M. J., Bunting, M. F., Hambrick, D. Z., Wilhelm, O., & Engle, R. W. (2005). Working memory span tasks: A methodological review and user's guide. *Psychonomic Bulletin & Review*, *12*, 769–786.
- D'Esposito, M., Postle, B. R., Jonides, J., & Smith, E. E. (1999). The neural substrate and temporal dynamics of interference effects in working memory as revealed by event-related functional MRI. *Proceedings of the National Academy of Sciences, USA*, *96*, 7514–7519.
- Dobbins, I. G., Rice, H. J., Wagner, A. D., & Schacter, D. L. (2003). Memory orientation and success: Separable neurocognitive components underlying episodic recognition. *Neuropsychologia*, *41*, 318–333.

- Dosher, B. A. (1976). The retrieval of sentences from memory: A speed-accuracy study. *Cognitive Psychology*, 8, 291-310.
- Dosher, B. A. (1981). The effects of delay and interference: A speed-accuracy study. *Cognitive Psychology*, 13, 551-582.
- Duarte, A., Ranganath, C., & Knight, R. T. (2005). Effects of unilateral prefrontal lesions on familiarity, recollection, and source memory. *Journal of Neuroscience*, 25, 8333-8337.
- Engle, R. W. (2002). Working memory capacity as executive attention. *Current Directions in Psychological Science*, 11, 19-23.
- Engle, R. W., Kane, M. J., & Tuholski, S. W. (1999). Individual differences in working memory capacity and what they tell us about controlled attention, general fluid intelligence, and functions of the prefrontal cortex. In A. Miyake & P. Shah (Eds.), *Models of working memory: Mechanisms of active maintenance and executive control* (pp. 102-134). London, England: Cambridge University Press.
- Ferredoes, E., Tononi, G., & Postle, B. R. (2006). Direct evidence for a prefrontal contribution to the control of proactive interference in verbal working memory. *Proceedings of the National Academy of Sciences, USA*, 103, 19530-19534.
- Foraker, S., & McElree, B. (2007). The role of prominence in pronoun resolution: Active versus passive representations. *Journal of Memory and Language*, 56, 357-383.
- Gronlund, S. D., Edwards, M. B., & Ohrt, D. D. (1997). Comparison of the retrieval of item versus spatial position information. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 23, 1261-1274.
- 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 and Language*, 33, 1-18.
- Jacoby, L. L. (1991). A process dissociation framework: Separating automatic from intentional uses of memory. *Journal of Memory and Language*, 30, 513-541.
- Jonides, J., Badre, D., Curtis, C., Thompson-Schill, S. L., & Smith, E. E. (2002). Mechanisms of conflict resolution in prefrontal cortex. In D. T. Stuss & R. T. Knight (Eds.), *The frontal lobes* (pp. 233-245). Oxford, England: Oxford University Press.
- Jonides, J., Marshuetz, C., Smith, E. E., Reuter-Lorenz, P. A., Koeppe, R. A., & Hartley, A. (2000). Age differences in behavior and PET activation reveal differences in interference resolution in verbal working memory. *Journal of Cognitive Neuroscience*, 12, 188-196.
- Jonides, J., & Nee, D. E. (2006). Brain mechanisms of proactive interference in working memory. *Neuroscience*, 139, 181-193.
- Jonides, J., Smith, E. E., Marshuetz, C., Koeppe, R. A., & Reuter-Lorenz, P. A. (1998). Inhibition in verbal working memory revealed by brain activation. *Proceedings of the National Academy of Sciences, USA*, 95, 8410-8413.
- Kane, M. J., Bleckley, M. K., Conway, A. R. A., & Engle, R. W. (2001). A controlled-attention view of working-memory capacity. *Journal of Experimental Psychology: General*, 130, 169-183.
- Kane, M. J., & Engle, R. W. (2000). Working-memory capacity, proactive interference, and divided attention: Limits on long-term memory retrieval. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 26, 336-358.
- Kane, M. J., & Engle, R. W. (2002). The role of prefrontal cortex in working-memory capacity, executive attention, and general fluid intelligence: An individual-differences perspective. *Psychonomic Bulletin & Review*, 9, 637-671.
- Kane, M. J., & Engle, R. W. (2003). Working-memory capacity and the control of attention: The contributions of goal neglect, response competition, and task set to Stroop interference. *Journal of Experimental Psychology: General*, 132, 47-70.
- Martin, A. E., & McElree, B. (2008). A content-addressable pointer mechanism underlies comprehension of verb-phrase ellipsis. *Journal of Memory and Language*, 58, 879-906.
- McElree, B. (1996). Accessing short-term memory with semantic and phonological information: A time-course analysis. *Memory & Cognition*, 24, 173-187.
- McElree, B. (1998). Attended and non-attended states in working-memory: Accessing categorized structures. *Journal of Memory and Language*, 38, 225-252.
- McElree, B. (2001). Working memory and focal attention. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 27, 817-835.
- McElree, B. (2006). Accessing recent events. In B. H. Ross (Ed.), *The Psychology of Learning and Motivation: Vol. 46. Advances in research and theory* (pp. 155-200). San Diego, CA: Academic Press.
- McElree, B., & Carrasco, M. (1999). The temporal dynamics of visual search: Evidence for parallel processing in feature and conjunctive searches. *Journal of Experimental Psychology: Human Perception and Performance*, 25, 1517-1539.
- McElree, B., Dolan, P. O., & Jacoby, L. L. (1999). Isolating the contributions of familiarity and source information in item recognition: A time course analysis. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 25, 563-582.
- McElree, B., & Dosher, B. A. (1989). Serial position and set size in short-term memory: The time course of recognition. *Journal of Experimental Psychology: General*, 18, 346-373.
- McElree, B., & Dosher, B. A. (1993). Serial retrieval processes in the recovery of order information. *Journal of Experimental Psychology: General*, 122, 291-315.
- McElree, B., Foraker, S., & Dyer, L. (2003). Memory structures that subserve sentence comprehension. *Journal of Memory and Language*, 48, 67-91.
- Monsell, S. (1978). Recency, immediate recognition memory, and reaction time. *Cognitive Psychology*, 10, 465-501.
- Murdock, B. B. (1971). A parallel-processing model for scanning. *Perception & Psychophysics*, 10, 289-291.
- Nelson, J. K., Reuter-Lorenz, P. A., Sylvester, C.-Y. C., Jonides, J., & Smith, E. E. (2003). Dissociable neural mechanisms underlying response-based and familiarity-based conflict in working memory. *Proceedings of the National Academy of Sciences, USA*, 100, 11171-11175.
- Oberauer, K. (2005). Binding and inhibition in working memory: Individual and age differences in short-term recognition. *Journal of Experimental Psychology: General*, 134, 368-387.
- Öztekin, I., Curtis, C. E., & McElree, B. (2009). The medial temporal lobe and the left inferior frontal gyrus jointly support interference resolution in verbal working memory. *Journal of Cognitive Neuroscience*, 21, 1967-1979.
- Öztekin, I., & McElree, B. (2007). Proactive interference slows recognition by eliminating fast assessments of familiarity. *Journal of Memory and Language*, 57, 126-149.
- Öztekin, I., McElree, B., Staresina, B. P., & Davachi, L. (2008). Working memory retrieval: Contributions of the left prefrontal cortex, the left posterior parietal cortex, and the hippocampus. *Journal of Cognitive Neuroscience*, 21, 581-593.
- Ratcliff, R. (1978). A theory of memory retrieval. *Psychological Review*, 85, 59-108.
- Ratcliff, R. (1980). A note on modeling accumulation of information when the rate of accumulation changes over time. *Journal of Mathematical Psychology*, 21, 178-184.
- Reed, A. (1973, August 10). Speed-accuracy trade-off in recognition memory. *Science*, 181, 574-576.
- Rosen, V. M., & Engle, R. W. (1997). The role of working memory capacity in retrieval. *Journal of Experimental Psychology: General*, 126, 211-227.

- Rouder, J. N., Speckman, P. L., Sun, D., Morey, R. D., & Iverson, G. (2009). Bayesian *t* tests for accepting and rejecting the null hypothesis. *Psychonomic Bulletin & Review*, *16*, 225–237.
- Snodgrass, J. G., & Corwin, J. (1988). Pragmatics of measuring recognition memory: Applications to dementia and amnesia. *Journal of Experimental Psychology: General*, *117*, 34–50.
- Thompson-Schill, S. L., Jonides, J., Marshuetz, C., Smith, E. E., D'Esposito, M., Kan, I. P., . . . Swick, D. (2002). Effects of frontal lobe damage on interference effects in working memory. *Cognitive, Affective, & Behavioral Neuroscience*, *2*, 109–120.
- Unsworth, N., & Engle, R. W. (2007). The nature of individual differences in working memory capacity: Active maintenance in primary memory and controlled search from secondary memory. *Psychological Review*, *114*, 104–132.
- Unsworth, N., Heitz, R. P., Schrock, J. C., & Engle, R. W. (2005). An automated version of the operation span task. *Behavior Research Methods*, *37*, 498–505.
- Van Overschelde, J. P., Rawson, K. A., & Dunlosky, J. (2004). Category norms: An updated and expanded version of the Battig and Montague (1969) norms. *Journal of Memory and Language*, *50*, 289–335.
- Wickelgren, W. A. (1977). Speed–accuracy tradeoff and information processing dynamics. *Acta Psychologica*, *41*, 67–85.
- Wickelgren, W. A., & Corbett, A. T. (1977). Associative interference and retrieval dynamics in yes–no recall and recognition. *Journal of Experimental Psychology: Human Learning and Memory*, *3*, 189–202.
- Wickelgren, W. A., Corbett, A. T., & Doshier, B. A. (1980). Priming and retrieval from short-term memory: A speed accuracy trade-off analysis. *Journal of Verbal Learning and Verbal Behavior*, *19*, 387–404.
- Yonelinas, A. P. (2002). The nature of recollection and familiarity: A review of 30 years of research. *Journal of Memory and Language*, *46*, 441–517.
- Zeelenberg, R., Wagenmakers, E.-J., & Shiffrin, R. M. (2004). Nonword repetition priming in lexical decision reverses as a function of study task and speed stress. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *30*, 270–277.

Received May 29, 2009

Revision received October 1, 2009

Accepted October 6, 2009 ■