Visual memories are stored along a logarithmically-compressed representation of the past

Inder Singh
Center for Memory and Brain
Department of Psychological and Brain Sciences
Boston University

Aude Oliva
Computer Science and Artificial Intelligence Laboratory
Massachussetts Institute of Technology

Marc W. Howard
Center for Memory and Brain
Department of Psychological and Brain Sciences
Boston University

Abstract

The recency effect is perhaps the most robust and important aspect of human memory. In continuous recognition the recency effect manifests as a decrease in accuracy and an increase in response time (RT) with the lag of a repeated stimulus. We used visual recognition with highly-memorable pictures to mitigate changes in accuracy and enable a detailed examination of the effect of recency on retrieval dynamics. The recency at which pictures were repeated ranged over more than two orders of magnitude from lag 1 up to lag 128. Analysis of the RT distributions showed that the time at which memories became accessible changed with the recency of the probe item. Additional analyses revealed that this effect was not attributable to an effect of immediate repetitions nor to increased processing fluency of the probe. These results suggest that visual memories can be accessed by sequentially scanning along a logarithmically-compressed representation of the past.

In recognition memory experiments participants must determine whether a probe stimulus has been previously experienced or not. As the recency of a repeated probe de-

The complete set of images can be downloaded from `http://cvcl.mit.edu/MM/stimuli.html`. Address correspondence to Marc Howard, Center for Memory and Brain, Department of Psychological and Brain Sciences, 2 Cummington Mall, Boston, MA 02215, marc777@bu.edu. Supported by NSF BCS-1058937 and AFOSR FA9550-12-1-0369 to MWH and a Google award to AO.

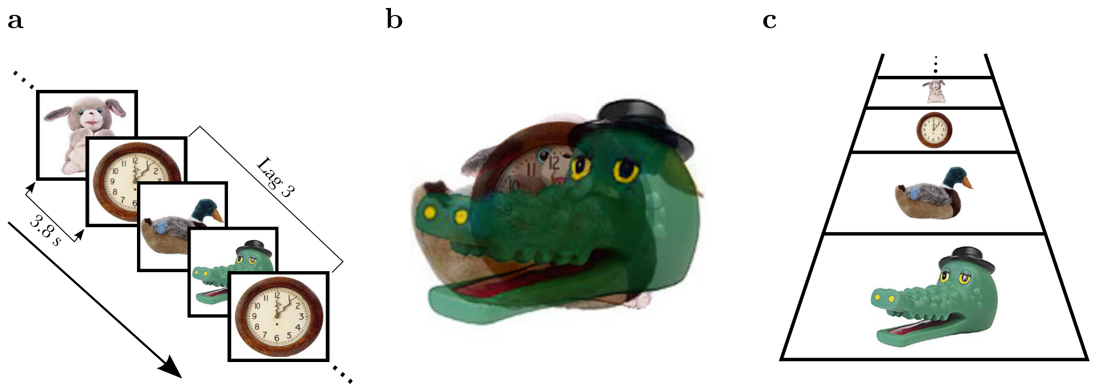

Figure 1. Accessing a memory representation to support recognition of visual stimuli. **a.** Schematic of a continuous recognition experiment. Participants experience a long sequence of pictures. Their task is to detect occasional repeated pictures. The variable "lag" measures the difference in position between the two presentations of the repeated picture. **b.** Cartoon representing a composite memory memory representation where more recent items (alligator) are more prominent than less recent items (clock). **c.** Cartoon representing a logarithmically-compressed representation. The representation contains information about the order in which items were experienced. Because it is ordered this representation that can be serially scanned starting at the present (alligator) and going back through the past. The foreshortening of the image is intended to suggest logarithmic compression of the representation.

creases, the accuracy of the judgment decreases and response time (RT) increases (Donkin & Nosofsky, 2012; Hockley, 1982; Monsell, 1978; Murdock & Anderson, 1975; Shepard & Teghtsoonian, 1961). Despite decades of empirical and modeling work, it remains unclear what changes in the state of memory cause the recency effect in recognition memory.

In continuous recognition, an item is presented at each time step; the participant indicates whether it was previously experienced. Because previously-experienced items must be identified from a stream of information, continuous recognition is somewhat similar to the experience of memory in the real world. Consider the task of an individual engaged in continuous recognition (Figure 1a). In order to correctly identify an item as old, one must compare it to the contents of their memory. In continuous recognition, the recency effect manifests as an nonlinear increase in RT with increasing lag of the repeated probe (Hintzman, 1969; Hockley, 1982; Okada, 1971); a very careful study by (Hockley, 1982) found a logarithmic increase in RT with increasing lag. Why does it take longer to retrieve memories from further in the past?

Many distributed memory models assume that memory is a composite store containing a noisy record of features from all the studied items (e.g., Anderson, 1973; Murdock, 1982; Shiffrin, Ratcliff, Murnane, & Nobel, 1993). A composite memory store can account for the recency effect if the features of items experienced further in the past are stored with less fidelity than items experienced more recently (Figure 1b). In contrast to the "bag of features" of a composite memory, another class of models proposes that features are stored along an ordered representation of experience (Figure 1c; G. D. A. Brown, Neath, & Chater, 2007; Murdock, 1974; Howard, Shankar, Aue, & Criss, 2015). As an analogy, the memory store behaves like a conveyor belt that recedes into the past (Murdock, 1974).

As each item is presented, it is placed at the front of the belt; previously-stored items shift back towards the past.[1] Both frameworks can handle the logarithmic increase in RTs. A composite memory can account for the logarithmic inrease if the strength of the match between a probe and the contents of memory decreases appropriately and this strength is coupled with a model of information accumulation (e.g., S. Brown & Heathcote, 2005; Ratcliff, 1978; Usher & McClelland, 2001). Information along an ordered representation can be sequentially accessed during memory search, which terminates when a match to the probe is found. Critically, if this ordered representation is compressed into a logarithmic scale (G. D. A. Brown et al., 2007; Chater & Brown, 2008; Shankar & Howard, 2013; Howard et al., 2015), then a logarithmic increase in RT with lag naturally results. On a logarithmic scale the difference between lag 1 and lag 2 is larger than the difference between lag 100 and lag 101. Rather, the difference between 1 and 2 is equivalent to the difference between 100 and 200.

While these two models cannot be distinguished based on RT alone, they make very different predictions regarding the *shape* of the RT distributions (Figure 2). Because all of the traces are stored together in a single composite representation (Figure 2a), the time needed to access a composite memory should be the same regardless of how far in the past the probe was experienced. However, the strength of that match should depend on the probe's recency. This is analogous to changing the drift rate in a drift diffusion process with increasing lag (Donkin & Nosofsky, 2012; Ratcliff, 1978). A composite memory representation suggests a parallel access model in which the RT distributions rise from zero at the same time but differ systematically in the tail of the distribution as a function of recency. If memories are aligned along an ordered representation very different qualitative predictions are possible. If the ordered representation is accessed via a serial scan that terminates when a match to the probe stimulus is found, then the time it takes to access the right memory should depend on the recency of the probe stimulus. Thus this kind of sequential access model results in RT distributions that start at different times (Figure 2b). If the strength of the match after the appropriate memory is accessed does not depend on recency, then the shape of the RT distribution should be unaffected by recency.

So which of these two hypotheses about the nature of memory representations best explains the recency effect? A parallel access one where all distributions rise at the same time with different rates resulting in different tails (Figure 2a) or a serial access one where distributions are simply shifted according to their recency (Figure 2b)? Unfortunately, the debate has, to our knowledge, not been resolved. The main issue is that as lag increases in continuous recognition, accuracy decreases (Hockley, 1982; Shepard & Teghtsoonian, 1961), making it more difficult to measure the effect of recency on retrieval dynamics independently of changes in accuracy. Brady, Konkle, Alvarez, and Oliva (2008) showed participants hundreds of memorable images in a continuous recognition task with lags varying over more than two orders of magnitude, with lags from 1 (no intervening items) to 128. Because the pictures were highly memorable, there was little variability in accuracy, even at very long lags. In addition the RT data are minimally affected by sequential dependencies, which are known to affect RTs in recognition memory (Malmberg & Annis, 2012). Because of the use of highly memorable pictures, wide range of lags tested, and the elimination of sequential

[1] In this study time per se and number of intervening items are confounded so we will not attempt to disentangle them.

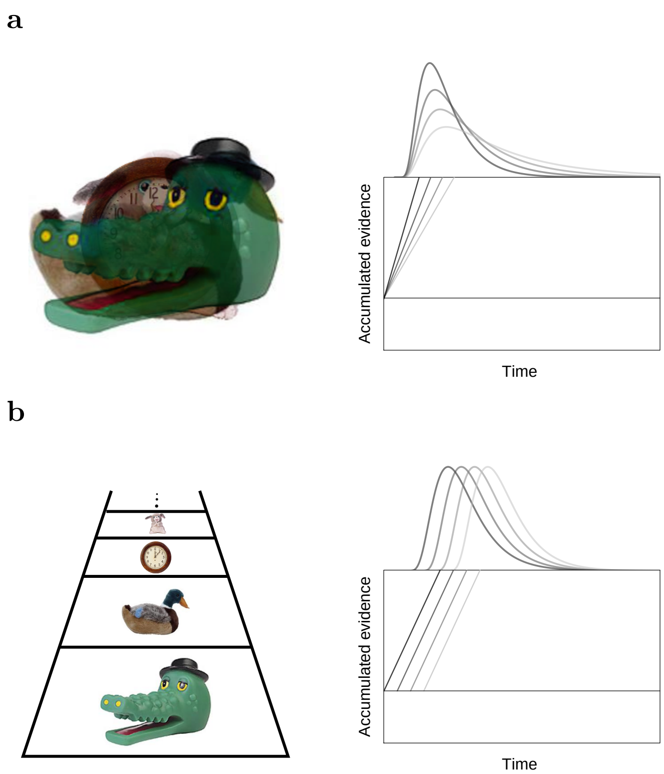

Figure 2. Two ways in which recency could affect response time distributions. **a.** A composite memory implies that the items are accessed in parallel. The rate of information accumulation is higher for more recent probes and the time to start accessing the memory of the probe item does not depend on its lag. Thus a parallel access hypothesis results in RT distributions that rise at about the same time but show systematically longer tails as the probe becomes less recent. **b.** A timeline can accommodate a self-terminating scanning hypothesis. Under this hypothesis, more recent probes do not differ in their rate of information accumulation, but rather the time that information starts to accumulate. Thus a scanning hypothesis results in RT distributions that rise at different times but maintain the same shape.

dependencies, the Brady et al. (2008) is well-suited to study the effect of recency on RT distributions.

Materials and Methods

Experiment

We analyzed the data collected as a part of the repeat detection task in the visual long term memory experiment conducted by Brady et al. (2008). During this task a total of 2896 images (2,500 new and 396 repeated images) were shown to 14 participants across 10 study blocks of approximately 20 minutes each. Categorically distinct images were obtained from a commercially available database (Hemera Photo-Objects,Vol.I and II) and through internet searches using Google Image Search. Examples of the images are used in Figure 1. The images (subtending 7.5° by 7.5° of visual angle) were presented for 3 s each, followed by an 800 ms fixation cross. Participants were required to respond to repeated images but not required to respond "no" to new items. The sequence of presentations did not include successive repetitions. As a result, times sequential responses were only made when a repetition was followed by a false alarm or vice versa. Because the false alarm rate was very low, this was quite rare, with successive responses on occurring a total of 3 times across all participants on the trials included in the analyses.

Lag was defined as the difference between the position of a repetition and the previous presentation of that stimulus; immediate repetition thus corresponds to a lag of 1. We analyzed memory for repeated pictures that were presented within the same block at lags from 1 to 128.[2] Small values of lag correspond to more recent probes, such that a recency effect is manifest as a decrease in accuracy or an increase in RT with increasing lag. In order to control for possible confounds of within-block position and lag, we discarded the first 128 images from each block. The way the experiment was constructed, repetitions at longer lags were somewhat less likely than repetitions at shorter lags. The average number of observations per participant included in our main analyses ranged from 27 at lag 1 to 9 at lag 128. Some of the images were repeated multiple times during the course of the experiment. Only the first repetitions are included in the first round of analyses. Second repetitions are analyzed later in the subsection entitled "Analysis of second repetitions."

Model fitting

In addition to standard distributional measures, we also characterized RT distributions using the shifted Wald distribution.[3] The shifted Wald distribution gives the finishing times for a drift diffusion process (Ratcliff, 1978; Donkin & Nosofsky, 2012) with one absorbing boundary. This is appropriate for this dataset because the participants only provided "yes" responses. The shifted Wald distribution is described by three parameters, μ, λ, and σ and is given by:

$$f(t;\mu,\lambda,\sigma) = \left[\frac{\lambda}{2\pi\,(t-\sigma)}\right] \exp \frac{-\lambda\,(t-\sigma-\mu)^2}{2\mu^2\,(t-\sigma)} \tag{1}$$

The parameter μ describes the rate of information accumulation, i.e., the drift rate. The parameter λ describes the distance of the boundary from the starting point of the diffusion process and σ describes the non-decision time before information begins to accumulate.

To determine if the change in RT with lag was due to changes in the rate of information accumulation or in the time at which information accumulation begins, we considered three models. In the *shift-only* model only σ was allowed to change as a function of lag. In the *drift-only* model only μ was allowed to change as a function of lag. In the *shift-plus-drift* model both σ and μ were allowed to vary freely as a function of lag. Note that both the shift and drift model are nested within the shift-plus-drift model.[4] We constrained all three parameters to be positive.

[2] We excluded lag 256, which had fewer than 5 responses per participant

[3] We found convergent results using the ex-Gaussian distribution, which we report in the supplementary information.

[4] Note that it is not sensible to allow boundary separation to vary as a function of recency—this would require that the participant know the recency of the probe before information begins to accumulate. An analogous hypothesis is tested in the context of the ex-gaussian distribution in the supplemental information.

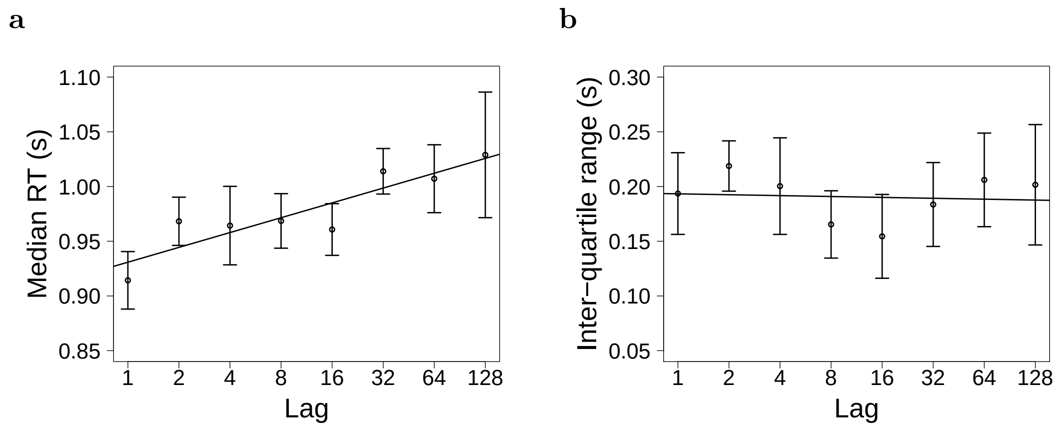

Figure 3. **a.** Median response time as a function of lag on $\log_2$ paper. Median response time increased approximately linearly with the logarithm of lag. The regression shows a slope of about 13 ms for each doubling of lag. **b.** The inter-quartile range as a function of lag on $\log_2$ paper. The regression slope is not reliably different from zero. Error bars in both panels represent the 95% confidence interval of the means normalized using the method described in Morey (2008).

Log likelihood of each response was computed using the analytic expression in Eq. 1 for each participant assuming responses are independent. The optimization function tried to minimize the negative log likelihood of the data given the parameters varying for each model using the Nelder-Mead algorithm. To avoid local minima, each parameter optimization was run from multiple starting points and the parameters from one iteration were passed back to the optimizer until the parameter values stopped changing. To compare the shift-model and the drift-model to the shift-plus-drift model we used the Bayesian Information Criterion.

Results

Accuracy was high at all lags

Participants showed very high accuracy for recognition of repeated images. Overall hit rate was .96, compared to a false alarm rate (incorrect detection of new images as repeated) of .01, corresponding to a d' of about 4. Hit rate decreased with lag, but remained quite high even for the longest lags. Even at lag 128, the hit rate was still .89 corresponding to a d' of about 3.5.

Non-parametric measures of RT distributions showed shifted RT distributions

We failed to observe any significant effect of block number (slope of reaction time with block was -0.002 ± 0.002, $t(125) = -0.56$). The median response time increased as a function of log lag (Figure 3a). Allowing for independent intercepts for each participant using mixed effects, we found that $\log_2$ lag was a significant predictor of the median response time, $.013 \pm .002$ (mean $\pm$ standard error), $t(97) = 5.82$, $p < 0.001$. The value of the regression coefficient indicates that each doubling [5] of the lag resulted in an increase of

[5]While the numerical value of lag is doubled, the elapsed time *per se* is not exactly doubled due to the ISI. There are various recognition studies in the literature showing interference and not time between stimuli

about 13 ms in the median RT. The total difference in the RT between lag 1 and lag 128 was on the order of 100 ms.

There is strong evidence suggesting that the time to access immediate repetitions is very different from the time to access repetitions after intervening stimuli, (e.g., McElree & Dosher, 1989). Note that with the definition of lag used here immediate repetitions have lag 1. [6] In order to ensure that the findings were not driven by immediate repetitions, we repeated the above analyses excluding lag 1. The slope of the median RT remained significant and not reliably different from the slope without immediate repetitions, $.011 \pm .002$, $t(83) = 3.85$, $p < 0.001$. The supplemental information reports all of the remaining analyses, including the model-based analyses, with immediate repetitions excluded. The conclusions are not affected by excluding immediate repetitions.

In contrast to the change in the median RT, the interquartile range did not change systematically as a function of lag. Using a linear mixed effects model, we found the slope of the interquartile range to be $-.001 \pm 0.003$, $t(97) = -.28$, placing it well outside the confidence interval of the slope for median RT. The visual appearance of the graph for interquartile range suggests that there might be a quadratic effect of log lag on the interquartile range. We further tested another mixed effects model that included a quadratic term and found no significant effects of either the quadratic $-.003 \pm 0.001$, $t(96) = -1.84$ or the linear component $-.019 \pm 0.010$, $t(96) = -1.84$ The lack of an effect of recency on interquartile range is consistent with the predictions of a serial self-terminating scanning model. In contrast, if recency affected drift rate, the interquartile range would be expected to grow with the median as the distributions become increasingly skewed.

Model-based analyses of RT distributions showed shifted RT distributions

The model-based analyses of RT distributions confirmed the impression from non-parametric statistics that the RT distributions shifted but did not systematically change shape with increasing lag. Table 1 shows the BIC values for the shift-only, drift-only, and shift-plus-drift models. The shift-only model, which only allowed the RT distributions to shift across lags provided a better fit than the drift-only model, which allowed only the rate of information accumulation to change across lags, for 13/14 participants $\chi^2(1) = 8.6$, $p < .01$. Even for the one participant for which the drift-only model provided a better fit to the data, the evidence in favor of the drift-only model was not overwhelming with the likelihood of the drift-only model only being about 4.4 times higher than the shift-only model. For comparison, the shift-only model was more than 20 times more likely than the drift-only model for 5/14 participants.

In addition, the shift-only model provided a better fit to the data than the shift-plus-drift model, in which both parameters were allowed to vary, for all 14 participants. Moreover, when both σ and μ were allowed to freely vary across conditions in the shift-plus-drift model, only σ changed systematically across lags. Treating σ as a dependent variable in a linear mixed effect model with $\log_2(\text{lag})$ as a regressor, we found a regression coefficient of $.019 \pm .003$, $t(97) = 6.18$, $p < .01$, corresponding to increase in the time to access memory of about 19 ms for each doubling of lag. To assess the effect of lag on μ in

accounts for the decrease in memory. (Kahana & Sekuler, 2002; Phillips & Christie, 1977).

[6] Other studies have defined lag such that immediate repetitions correspond to lag 0. However, Hockley (1982) found that RT was a function of the logarithm of lag as defined here.

	ΔBIC		
Participant	drift-only	shift-only	shift-plus-drift
1	16.76	*0*	20.88
2	0.85	*0*	33.99
3	2.53	*0*	16.27
4	7.71	*0*	7.10
5	0.46	*0*	29.79
6	8.60	*0*	30.87
7	12.00	*0*	27.63
8	*0*	0.96	24.31
9	3.81	*0*	11.98
10	25.34	*0*	25.79
11	0.36	*0*	28.99
12	0.21	*0*	28.39
13	19.77	*0*	2.27
14	4.24	*0*	14.43
Total	101.69	*0*	301.71

Table 1: Difference in BIC between a particular model and the best fitting model for each participant. Cases where the shift-only model provided a better fit than the shift-plus-drift model are set in *italics*.

similar units, we took the λ divided by μ[7] as a dependent measure. An analogous analysis showed a regression coefficient of $-.0003\pm.004$, $t(97) = -.08$. These values can be compared to the experimentally-observed changes in median RT. The change in σ overlapped with the experimentally-observed effect of lag whereas the change in μ overlapped with zero and was much smaller than the experimentally-observed effect. Thus even when both parameters were allowed to freely vary, the experimentally-observed change in RT as a function of lag was carried exclusively by changes in σ.

Analysis of second repetitions argued against a non-scanning account of the results

We have shown strong evidence that the effect of recency on RT distributions is solely attributable to a shift in the distribution, consistent with a change in non-decision time in the drift diffusion model. While a shift in the RT distribution is consistent with a serial self-terminating search during the memory comparison phase, this is not the only possible explanation. Presumably, the probe must be encoded before it can be compared to memory. The shift in the RT distributions is also consistent with the hypothesis that encoding of recently-experienced probes is facilitated but that there is no effect of recency on the memory-comparison stage *per se* (Sternberg, 1969).

If the recency of a repeated item allows it to be processed faster as a probe, then repeating the item again should have an additional effect on RT. Thus far we have examined RTs to the first repetition (second presentation) of the probe stimulus. In order to evaluate the hypothesis that the recency effect was attributable to processing fluency, we examined

[7]The drift rate μ is in units of evidence per unit time. The boundary separation λ is in units of evidence. λ/μ thus has units of time, making it directly comparable to σ.

RTs to the second repetition (third overall presentation). If changes in RTs were being driven by greater facility of processing of recently-presented probes, then the lags of the two presentations ought to both affect RT. In contrast, if the changes in RT are driven by a self-terminating scanning model during the memory comparison stage, then only the most recent lag should affect RT.

All participants correctly identified all 45 of the 45 images that were repeated a second time as having been previously presented. We compared the RTs for the second repetition to the RTs for the first repetition of the same images. We found no significant difference between the RT distributions of first and second repetitions (two-sided Kolmogorov-Smirnov test, $D = 0.067$). The median of the first repetition was 0.96 ± 0.04 and the median for the second repetition was 0.93 ± 0.04. The foregoing results indicate that correct RTs to second presentations were not faster than RTs to third presentations, inconsistent with the processing fluency account. To determine if RT on the third presentation was consistent with serial scanning, we examined the effect of recency on the RT to the second repetition. There are two lags associated with the third presentation of a probe. Let us refer to the lag between the first and second presentation as lag_1 and the lag between the second and third presentation as lag_2, so that at the third presentation lag_2 is the most recent lag. Allowing each participant to have an independent intercept (to account for between participant variability) in a linear mixed effects model using $\log_2$ of the two lags as regressors, we found a significant effect of the base 2 log of lag_2 of about the same magnitude as the effect of lag on the first repetition (Figure 3a), $.012 \pm .004$, $t(614) = 2.72$, $p < .01$. In contrast, there was not a significant additional effect of lag_1, $.001 \pm .003$, $t(614) = .45$. To summarize, while there was an effect of the lag to the most recent presentation of the probe stimulus, there was no additional effect of its prior presentation. These observations fail to provide any evidence for the hypothesis that the effect of recency on RT was due to a facilitation of probe encoding. In contrast, these findings are exactly as one would have predicted if the effect of recency on RT was caused by serial self-terminating scanning of a logarithmically-compressed representation of the past.

Discussion

RT distributions shifted by about 13 ms for each doubling of lag. There was no sign of a change in the shape of the RT distributions with lag, as if RT depended on scanning a logarithmically-compressed ordered representation of the past. These conclusions are unlikely to depend critically on the choice of the shifted Wald distribution to characterize the distributions. All models of RT can include a non-decision component (e.g., S. Brown & Heathcote, 2005; Ratcliff, 1978; Usher & McClelland, 2001) that simply shifts the observed distributions. We also reached the same conclusions using the ex-Gaussian distribution (see Supplemental Information). Moreover, the interquartile range, a purely descriptive measure of the RT distributions, was unaffected by the lag of probes (Figure 3b). In the context of widely-used information accumulation models of RT (S. Brown, Steyvers, & Hemmer, 2007; Ratcliff, 1978), there is little doubt that these data show evidence that the recency effect is driven by time to retrieve a memory rather than the rate of information accumulation. Of course, it is possible that these models are incorrect in a basic way. For instance, perhaps drift rate changes over time in some complicated way. Perhaps it is possible to account for what appears to be a straightforward shift in the distributions with a complex set of assumptions about the trial-to-trial variability in drift rate rather that mimics a simpler shift. However, within the context of simple information-accumulation models, the result seems unambiguous.

We did not find any evidence to support the hypothesis that the shift in RT distributions with recency was attributable to a facilitation of probe encoding. Instead, our results are consistent with scanning of an ordered representation of the past with logarithmic spacing. The range of lags suggests that the logarithmic timeline changes over two orders of magnitude from a few seconds (lag 1 corresponds to 3.8 s) up to more than eight minutes (lag 128 corresponds to 486 s). The fact that none of our findings were attributable to immediate repetitions suggests a continuity of this representation over a relatively large interval.

Serial scanning of an ordered representation requires a much more elaborate memory representation than a simple composite memory. However, it also suggests a deep analogy between search through memory and perceptual attention. We can deploy attention strategically to sensory dimensions resulting in preferential access to information available along those dimensions (e.g., Teder-Sälejärvi, Münte, Sperlich, & Hillyard, 1999; Shomstein & Yantis, 2004). For instance, directing attention to a particular part of visual space enhances one's ability to detect stimuli presented at that location (e.g., Posner, Snyder, & Davidson, 1980). A logarithmic scale would naturally give rise to the Weber-Fechner law. Starting from the perspective of perception, the only assumption needed to implement the scanning model is that the past is represented like a perceptual dimension (G. D. A. Brown et al., 2007; Howard et al., 2015).

Many previous studies have found evidence for parallel access to memory, not sequential scanning, in study-test recognition (Nosofsky, Little, Donkin, & Fific, 2011; Ratcliff & Murdock, 1976; McElree & Dosher, 1989; Hockley, 1984; Nosofsky, Cox, Cao, & Shiffrin, 2014). Moreover the finding that RTs were not different on the second repetition of a probe is not consistent with previous studies (e.g., Hockley, 1982; Hintzman, 1969). Several potentially important methodological differences may account for thes discrepancies. This experiment used continuous recognition rather than the study-test procedure (Nosofsky et al., 2011; Ratcliff & Murdock, 1976; McElree & Dosher, 1989; Hockley, 1984; Nosofsky et al., 2014) and highly memorable trial-unique visual stimuli. In addition, this study only required positive responses to repeated stimuli. Combined with the structure of the test sequence, sequential dependencies were likely to have less effect than in the overwhelming majority of recognition studies. The question of which combination of these methodological differences accounts for the evidence for serial scanning is an extremely important one that merits further investigation. It is worth noting that an ordered representation could also in principle be accessed in parallel (Howard et al., 2015), whereas it is not clear how (or why) a composite representation could be scanned in a recognition memory task.

References

Anderson, J. A. (1973). A theory for the recognition of items from short memorized lists. *Psychological Review*, *80*, 417-438.

Brady, T. F., Konkle, T., Alvarez, G. A., & Oliva, A. (2008). Visual long-term memory has a massive storage capacity for object details. *Proceedings of the National Academy of Sciences*, *105*(38), 14325–14329.

Brown, G. D. A., Neath, I., & Chater, N. (2007). A temporal ratio model of memory. *Psychological Review*, *114*(3), 539-76.

Brown, S., & Heathcote, A. (2005). A ballistic model of choice response time. *Psychological Review*, *112*(1), 117-28. doi: 10.1037/0033-295X.112.1.117

Brown, S., Steyvers, M., & Hemmer, P. (2007). Modeling experimentally induced strategy shifts. *Psychological Science*, *18*(1), 40-5.

Chater, N., & Brown, G. D. A. (2008). From universal laws of cognition to specific cognitive models. *Cognitive Science*, *32*(1), 36-67. doi: 10.1080/03640210701801941

Donkin, C., & Nosofsky, R. M. (2012). A power-law model of psychological memory strength in short- and long-term recognition. *Psychological Science*. doi: 10.1177/0956797611430961

Hintzman, D. L. (1969). Recognition time: Effects of recency, frequency and the spacing of repetitions. *Journal of Experimental Psychology*, *79*(1p1), 192.

Hockley, W. E. (1982). Retrieval processes in continuous recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *8*, 497-512.

Hockley, W. E. (1984). Analysis of response time distributions in the study of cognitive processes. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *10*(4), 598-615.

Howard, M. W., Shankar, K. H., Aue, W., & Criss, A. H. (2015). A distributed representation of internal time. *Psychological Review*, *122*(1), 24-53.

Howard, M. W., Shankar, K. H., Aue, W., & Criss, A. H. (in press). A distributed representation of internal time. *Psychological Review*.

Kahana, M. J., & Sekuler, R. (2002). Recognizing spatial patterns: a noisy exemplar approach. *Vision Research*, *42*(18), 2177-192.

Malmberg, K. J., & Annis, J. (2012). On the relationship between memory and perception: Sequential dependencies in recognition memory testing. *J Exp Psychol Gen*, *141*(2), 233-59. doi: 10.1037/a0025277

Matzke, D., & Wagenmakers, E.-J. (2009). Psychological interpretation of the ex-gaussian and shifted wald parameters: a diffusion model analysis. *Psychonomic Bulletin & Review*, *16*(5), 798-817. doi: 10.3758/PBR.16.5.798

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*, *118*, 346-373.

Monsell, S. (1978). Recency, immediate recognition memory, and reaction time. *Cognitive Psychology*, *10*, 465-501.

Morey, R. D. (2008). Confidence intervals from normalized data: A correction to Cousineau (2005). *Reason*, *4*(2), 61–64.

Murdock, B. B. (1974). *Human memory: Theory and data*. Potomac, MD: Erlbaum.

Murdock, B. B. (1982). A theory for the storage and retrieval of item and associative information. *Psychological Review*, *89*, 609-626.

Murdock, B. B., & Anderson, R. E. (1975). Encoding, storage and retrieval of item information. In R. L. Solso (Ed.), *Information processing and cognition: The loyola symposium* (p. 145-194). Hillsdale, New Jersey: Erlbaum.

Nosofsky, R. M., Cox, G. E., Cao, R., & Shiffrin, R. M. (2014). An exemplar-familiarity model predicts short-term and long-term probe recognition across diverse forms of memory search.

Nosofsky, R. M., Little, D. R., Donkin, C., & Fific, M. (2011). Short-term memory scanning viewed as exemplar-based categorization. *Psychological Review*, *118*(2), 280-315.

Okada, R. (1971). Decision latencies in short-term recognition memory. *Journal of Experimental Psychology*, *90*(1), 27.

Phillips, W., & Christie, D. (1977). Components of visual memory. *The Quarterly Journal of Experimental Psychology*, *29*(1), 117–133.

Posner, M. I., Snyder, C. R., & Davidson, B. J. (1980). Attention and the detection of signals. *Journal of experimental psychology: General*, *109*(2), 160.

Ratcliff, R. (1978). A theory of memory retrieval. *Psychological Review*, *85*, 59-108.

Ratcliff, R., & Murdock, B. B. (1976). Retrieval processes in recognition memory. *Psychological Review*, *83*(3), 190-214.

Shankar, K. H., & Howard, M. W. (2013). Optimally fuzzy scale-free memory. *Journal of Machine Learning Research*, *14*, 3753-3780.

Shepard, R. N., & Teghtsoonian, M. (1961). Retention of information under conditions approaching a steady state. *Journal of Experimental Psychology*, *62*, 302-309.

Shiffrin, R. M., Ratcliff, R., Murnane, K., & Nobel, P. (1993). Todam and the list-strength and list-length effects: A reply to Murdock and Kahana. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *19*, 1445-1449.

Shomstein, S., & Yantis, S. (2004). Control of attention shifts between vision and audition in human cortex. *The Journal of Neuroscience*, *24*(47), 10702–10706.

Sternberg, S. (1969). The discovery of processing stages: Extensions of Donders' method. *Acta Psychologica*, *30*, 276-315.

Teder-Sälejärvi, W. A., Münte, T. F., Sperlich, F.-J., & Hillyard, S. A. (1999). Intra-modal and cross-modal spatial attention to auditory and visual stimuli. an event-related brain potential study. *Cognitive Brain Research*, *8*(3), 327–343.

Usher, M., & McClelland, J. L. (2001). The time course of perceptual choice: the leaky, competing accumulator model. *Psychological Review*, *108*(3), 550-92.

Best fitting parameter values for shift-only model

Participant	μ	λ	$\sigma[1]$	$\sigma[2]$	$\sigma[4]$	$\sigma[8]$	$\sigma[16]$	$\sigma[32]$	$\sigma[64]$	$\sigma[128]$
1	3.034	1.185	0.572	0.648	0.769	0.705	0.718	0.748	0.761	0.721
2	5.991	5.296	0.010	0.038	0.066	0.074	0.037	0.120	0.029	0.097
3	3.342	1.415	0.481	0.491	0.474	0.586	0.556	0.547	0.568	0.598
4	5.257	2.121	0.588	0.530	0.568	0.560	0.484	0.565	0.607	0.510
5	7.608	7.091	0.052	0.025	0.017	0.010	0.073	0.043	0.078	0.070
6	2.053	0.953	0.543	0.558	0.644	0.698	0.645	0.713	0.611	0.654
7	3.591	1.400	0.537	0.642	0.641	0.652	0.668	0.693	0.754	0.767
8	2.556	1.806	0.392	0.412	0.406	0.439	0.479	0.490	0.510	0.685
9	2.484	0.932	0.513	0.481	0.597	0.597	0.581	0.717	0.628	0.674
10	2.750	0.530	0.435	0.483	0.496	0.517	0.566	0.585	0.565	0.653
11	6.143	5.428	0.058	0.097	0.104	0.129	0.060	0.179	0.152	0.137
12	4.938	1.704	0.663	0.718	0.727	0.691	0.703	0.718	0.698	0.795
13	3.710	0.674	0.445	0.538	0.574	0.551	0.544	0.631	0.597	0.601
14	4.309	3.005	0.503	0.523	0.603	0.533	0.678	0.686	0.675	0.774

Table 2: Subject-wise parameter fits for the Wald distribution under the shift-only model. For each subject, σ, corresponding to the non-decision time was allowed to change as a function of lag. See methods for details.

Supplemental Information

Subject-wise Wald distribution parameter values for the drift-only and shift-only models

The results section presents the overall goodness of fits and summary statistics for the parameter values from fitting a Wald distribution to the RT distributions. Table 2 and table 3 report the actual parameter values for the model fits for each subject for the *shift-only* and the *drift-only* models respectively.

Ex-Gaussian analyses of RT distributions

The results from the model-based analysis of the shifted Wald distribution showed that the shift-only model, which only allowed the shift parameter to vary with the lag provided a better fit to the RT distributions as compared to the drift-only model, which allowed only the rate of information accumulation to vary across lags. The same analysis is reported here using an ex-Gaussian distribution. The ex-Gaussian distribution has been used for decades to characterize the distributional properties of the reaction time distributions (e.g., Hockley, 1984; Ratcliff & Murdock, 1976).

The ex-Gaussian distribution results from the convolution of a Gaussian process and an exponential process. This results in a distribution with three parameters μ, σ[8] and τ. The μ and σ parameters characterize the mean and the standard deviation of the Gaussian; τ characterizes the mean of the exponential component. While the parameters of the ex-Gaussian can be difficult to interpret in terms of process models of decision-making (Matzke

[8] The"σ" used here for the ex-Gaussian distribution is different from the σ used earlier for the Wald distribution. In the Wald distribution, σ represents the "shift" or the non-decision time. For an ex-Gaussian distribution, σ represents the standard deviation of the Gaussian component.

				Best fitting parameter values for drift-only model					
Participant	$\mu[1]$	$\mu[2]$	$\mu[4]$	$\mu[8]$	$\mu[16]$	$\mu[32]$	$\mu[64]$	$\mu[128]$	λ	σ
1	4.242	4.353	3.922	4.479	4.130	3.915	3.618	3.778	2.574	0.456
2	6.459	6.156	6.047	6.008	6.261	5.739	6.254	5.747	5.673	0.010
3	4.536	4.133	3.876	3.903	3.583	3.225	4.043	3.915	2.084	0.420
4	6.204	6.096	6.182	6.029	6.640	6.485	5.912	7.139	3.506	0.396
5	7.739	7.871	7.944	8.037	7.480	7.597	7.529	7.556	7.472	0.010
6	3.224	2.965	2.263	2.397	2.622	1.991	2.848	2.075	1.549	0.486
7	5.534	4.544	4.908	4.220	4.641	4.229	4.140	3.793	2.660	0.457
8	3.187	2.873	3.510	3.167	3.039	2.645	2.489	2.494	2.524	0.301
9	4.216	3.430	3.716	3.913	4.035	3.023	2.472	3.603	2.035	0.383
10	6.095	4.209	4.518	4.847	3.724	3.327	3.851	2.186	1.315	0.393
11	6.862	6.492	6.439	6.340	6.938	6.141	6.172	6.371	6.372	0.010
12	6.299	5.020	4.806	5.471	5.431	5.080	5.225	4.613	2.042	0.665
13	7.697	6.770	6.426	6.750	7.093	5.297	4.306	5.220	2.476	0.333
14	5.625	5.711	5.061	5.576	5.106	4.945	5.029	4.072	5.389	0.265

Table 3: Subject-wise parameter fits for the Wald distribution under the drift-only model. For each subject, μ, corresponding to the drift rate was was allowed to change as a function of lag. See methods for details.

& Wagenmakers, 2009), under some circumstances the model can yield straightforward conclusions. In particular, changing the μ parameter shifts the RT distribution to the right. Changes in non-decision time result solely in changes in μ.

Results. To assess the effects of recency on the parameters of the ex-Gaussian, we fit six models using methods identical to those used for the shifted Wald distribution. Table 4 shows the difference in BIC values for the μ-only, σ-only, τ-only, μ-plus-τ, σ-plus-τ, and μ-plus-σ-plus-τ models for each participant. This result is consistent with the results obtained using the shifted Wald distribution. We find again that 12/14 participant $\chi^2(1) = 5.8$, $p < .02$ show the μ-only model to be a better fit than the τ-only model. Further the μ-only model provided a better fit to the data than the μ-plus-τ for 13/14 participant and the μ-plus-σ, σ-plus-τ and μ-plus-σ-plus-τ models for 14/14 participants.

In terms of trends in parameter change by lag, when both μ and τ are allowed to vary in the μ-plus-τ model, the change was systematically carried by the μ parameter and not the τ parameter. Allowing each participant to have an independent intercept (to account for between subject variability) in a linear mixed effects model using $\log_2(\text{lag})$ as a regressor, we find the slope of μ parameter to be $.017 \pm .0024$, $t(97) = 6.82$, $p < .01$. The τ parameter in the μ-plus-τ model on the other hand has a not significant slope of $.003 \pm .003$, $t(97) = 0.77$. This is consistent with the results obtained using the shifted Wald distribution and the experimentally-observed changes in the median RT. Similarly, in the μ-plus-σ model, once again the μ systematically changes with $\log_2(\text{lag})$ with a slope of $.017 \pm .002$, $t(97) = 7.71$, $p < .01$ while the slope of the σ parameter is not significantly different from zero $-.001 \pm .003$, $t(97) = -0.24$, $p = .81$.

For completeness we also tested a model where all three parameters μ, σ and τ were

	ΔBIC						
Participant	μ-only	σ-only	τ-only	μ-plus-σ	μ-plus-τ	σ-plus-τ	μ-plus-σ-plus-τ
1	*1.64*	0	13.04	25.87	28.09	27.43	53.80
2	*0*	5.12	1.12	32.37	32.34	34.13	63.97
3	*0.43*	0	7.50	13.90	26.31	25.42	42.74
4	*2.71*	0	5.07	25.17	21.98	25.71	46.47
5	0.92	0	0.44	29.40	31.68	30.45	64.80
6	*0*	22.60	21.38	51.14	43.59	43.52	74.06
7	*0*	21.84	8.11	25.32	28.09	37.81	55.42
8	7.58	12.41	0	33.48	22.09	27.43	41.03
9	*0*	18.10	7.77	23.06	17.46	27.78	39.56
10	*0*	41.95	18.75	25.64	19.07	39.87	37.41
11	*0*	6.85	0.91	26.28	30.62	30.85	55.88
12	*0*	11.98	4.18	23.75	30.04	36.12	59.68
13	*3.72*	28.66	15.54	25.65	0	11.87	28.75
14	*0*	3.09	6.22	10.04	20.40	17.64	36.91
Total	*0*	155.60	93.03	354.05	334.74	399.02	683.46

Table 4: Model fitting using the ex-Gaussian distribution: Difference in BIC between a particular model and the best fitting model for each participant using the ex-Gaussian distribution. Cases where the μ-only model provided a better fit than the τ model are set in *italics*. All 14 subjects show a better fit for the μ-only as compared to the μ-plus-σ-plus-τ model

allowed to freely vary. Again the μ parameter systematically varied with $\log_2$(lag), $.018 \pm .003$, $t(97) = 5.5$, $p < .01$ whereas the slopes of σ and τ parameter were not significantly different from zero and much smaller than the range expected for the data. The slope of the σ parameter was $-.002 \pm .002$, $t(97) = -1.14$ and the slope of the τ parameter was $.001 \pm .003$, $t(97) = 0.28$. Thus when all the three parameters of the ex-Gaussian distribution are allowed to freely vary, the change in the μ parameter with lag corresponds to the observed change in the median RTs while there was no significant effect of lag on the σ and τ parameters.

Replication of results excluding immediate repetitions

There is strong evidence suggesting that the time to access immediate repetitions is very different from the time to access repetitions after intervening stimuli, (e.g., McElree & Dosher, 1989). Note that with the definition of lag used here immediate repetitions have lag 1.[9] In order to ensure that the findings were not driven by immediate repetitions, we repeated the analyses reported in the Results section excluding lag 1.

Non-parametric measures of RT distributions. The slope of the median RT with $\log_2$(lag) remained significant, and was similar to the slope observed with immediate repetitions included. $.011 \pm .002$, $t(83) = 3.85$, $p < 0.001$. The slope of the IQR was once again found to be not significantly different from zero, $.001 \pm .004$, $t(83) = -0.22$, $p = 0.829$.

[9]Other studies have defined lag such that immediate repetitions correspond to lag 0. However, Hockley (1982) found that RT was a function of the logarithm of lag as defined here.

Given the visual appearance of the IQR plot, we also tested whether there was a quadratic effect of $\log_2$(lag) on the IQR. We regressed the RT on the lag and the square of the lag, we found a significant effect for both terms. The slope of the square of the $\log_2$(lag) was $.005 \pm .002$, $t(82) = 2.59$, $p = 0.011$ and the linear component was $\log_2$(lag), is $-.042 \pm .016$, $t(82) = -2.58$, $p < 0.02$. Under a drift-only model, the IQR should monotonically increase with lag.

Analysis of second repetitions. In the Results section, we showed evidence from the analysis of multiple repetitions that argued against a non-scanning model. The experiment included trials where the probe were repeated a second time (third overall presentation). We found that there was no overall difference in the medians and a linear mixed effects model regressing the RT on the lag between the first and second presentation, lag_1 and the lag between the second and third presentation as lag_2 (the third presentation lag_2 is the most recent lag) found that the RT depended on the more recent lag, lag_2 .

Here we ran the same analyses excluding the trials where either the first or the second repetition was an immediate repetition (lag 1). For the remaining trials, the median of the first repetition was 0.98 ± 0.04 and the median for the second repetition was 0.93 ± 0.04. Allowing each participant to have an independent intercept (to account for between participant variability) in a linear mixed effects model using $\log_2$ of the two lags as regressors, we found a significant effect of the base 2 log of lag_2 of about the same magnitude as the effect of lag on the first repetition (Figure 3a), $.015 \pm .004$, $t(502) = 20.76$, $p < .01$. In contrast, there was no significant additional effect of lag_1, $.005 \pm .004$, $t(502) = 1.53$. This replicates the findings presented in the Results section even when we exclude immediate repetitions from the analysis.

Model-based analyses of RT distributions . The model-based analyses of RT distributions confirmed the impression from non-parametric statistics that the RT distributions shifted but did not systematically change shape with increasing lag even when we exclude immediate repetitions from the analysis. Table 5 shows the BIC values for the shift-only, drift-only, and shift-plus-drift models. The shift-only model, which only allowed the RT distributions to shift across lags provided a better fit than the drift-only model, which allowed only the rate of information accumulation to change across lags, for 11/14 participants. The ΔBIC values for drift-only model were significantly different from 0, $t(13) = 2.6441$, $p = 0.02$.

Even when immediate repetitions were excluded from the analysis, the shift-only model provided a better fit to the data than the shift-plus-drift model, in which both parameters were allowed to vary, for all 14 participants. Moreover, when both σ and μ were allowed to freely vary across conditions in the shift-plus-drift model, only σ changed systematically across lags. Treating σ as a dependent variable in a linear mixed effect model with $\log_2$(lag) as a regressor, we found a regression coefficient of $.020 \pm .004$, $t(83) = 5.51$, $p < .01$, corresponding to increase in the time to access memory of about 20 ms for each doubling of lag. To assess the effect of lag on μ in similar units, we took the λ divided by μ as a dependent measure. An analogous analysis showed a regression coefficient of $-.002 \pm .005$, $t(83) = -.32$. These values can be compared to the experimentally-observed changes in median RT. The change in σ is comparable to the experimentally-observed effect of lag whereas the change in μ overlapped with zero and was much smaller than the

		ΔBIC	
Participant	drift-only	shift-only	shift-plus-drift
1	14.29	*0*	25.15
2	0.60	*0*	29.40
3	1.34	*0*	12.37
4	4.77	*0*	11.73
5	0.25	*0*	26.84
6	7.15	*0*	26.67
7	11.10	*0*	23.46
8	0	*1.39*	20.89
9	0	*1.73*	7.93
10	24.62	*0*	23.19
11	0.47	*0*	24.11
12	2.58	*0*	25.94
13	0	*19.26*	20.61
14	2.28	*0*	11.00
Total	47.06	*0*	266.90

Table 5: Difference in BIC between a particular model and the best fitting model for each participant. Cases where the shift-only model provided a better fit than the shift-plus-drift model are set in *italics*. Overall the shift-only model prvides a better fit as compared to the drift-only and the shift-plus-drift model.

experimentally-observed effect. Thus even when both parameters were allowed to freely vary, the experimentally-observed change in RT as a function of lag was carried exclusively by changes in σ.