

University of Warwick institutional repository

This paper is made available online in accordance with publisher policies. Please scroll down to view the document itself. Please refer to the repository record for this item and our policy information available from the repository home page for further information.

To see the final version of this paper please visit the publisher's website. Access to the published version may require a subscription.

Author(s): Paolo Martini and Vera Maljkovic

Article Title: Short-term memory for pictures seen once or twice

Year of publication: 2009

Link to published version:

<http://dx.doi.org/10.1016/j.visres.2009.04.007>

Publisher statement: none

SHORT-TERM MEMORY FOR PICTURES SEEN ONCE OR TWICE

Paolo Martini and Vera Maljkovic

Department of Psychology, University of Warwick

Corresponding Author: Paolo Martini, p.martini@warwick.ac.uk

ABSTRACT

The present study is concerned with the effects of exposure time, repetition, spacing and lag on old/new recognition memory for generic visual scenes presented in a RSVP paradigm. Early memory studies with verbal material found that knowledge of total exposure time at study is sufficient to accurately predict memory performance at test (the Total Time Hypothesis), irrespective of number of repetitions, spacing or lag. However, other studies have disputed such simple dependence of memory strength on total study time, demonstrating superadditive facilitatory effects of spacing and lag, as well as inhibitory effects, such as the Ranschburg effect, Repetition Blindness and the Attentional Blink. In the experimental conditions of the present study we find no evidence of either facilitatory or inhibitory effects: recognition memory for pictures in RSVP supports the Total Time Hypothesis. The data are consistent with an Unequal Variance Signal Detection Theory model of memory that assumes the average strength and the variance of the familiarity of pictures both increase with total study time. The main conclusion is that the growth of visual scene familiarity with temporal exposure and repetition is a stochastically independent process.

ACKNOWLEDGMENTS

Address for correspondence: Paolo Martini, Department of Psychology, The University of Warwick, Gibbet Hill Road, Coventry, CV4 7AL, UK; e-mail: p.martini@warwick.ac.uk, v.maljkovic@warwick.ac.uk. We thank Gordon Brown and Luciano Buratto for discussions and Peter Chang, Julie Gabrielle Berger and Romy Powert for help with the experimental design and the running of the study. Supported by NIH grant R01-EY13155 to V. Maljkovic and RCUK fellowship to P.Martini.

INTRODUCTION

It is common knowledge that memory depends on time and practice. Repeating the material to be remembered and/or studying it for a longer time increases the probability of correctly recalling it at a later epoch (the Law of Practice) and quality of recall declines with the passage of time since last study (the Law of Recency). While these psychological “laws” of memory may seem evident even from casual observations, scientific investigations of temporal aspects of memory function, ongoing since the classical studies of Ebbinghaus (Ebbinghaus, 1964), have provided a more nuanced view of the influence of time and practice on various aspects of memory performance (Roediger, 2008). Consider the Law of Practice: early studies showed that repeated exposure to an item over two separate epochs each lasting a time T affords the same memory performance as exposure to the same item over a single epoch lasting a time $2T$ (Bugelski, 1962, Cooper & Pantle, 1967), suggesting that total exposure time completely predicts memory strength (the Total Time Hypothesis). However, later studies (Cepeda, Pashler, Vul, Wixted & Rohrer, 2006) have disputed such simple dependence of memory strength on total study time and have provided evidence for super-additive, facilitatory effects: it has been shown that under appropriate conditions there is greater improvement when repeated items are spaced (separated in time by intervening items or interruptions) rather than massed (the Spacing Effect), and that performance may improve further as a function of temporal distance between repetitions (the Lag Effect).

The present study is concerned with the effects of exposure time, repetition, spacing and lag on memory for generic visual scenes, such as those normally encountered in everyday life by exploring the visual environment. The range of temporal exposures that were investigated in this

project, from a few milliseconds to a few seconds, is representative of the normal range of durations of ocular fixations (Harris, Hainline, Abramov, Lemerise & et al., 1988). The number of pictures displayed at study (the list length) was limited to less than 10, affording close to perfect performance at the longest exposures used. An old-new recognition memory test was adopted and the test was completed within a period of about a minute from the beginning of the study phase. Thus, it may be said that the present study is concerned with what is customarily referred to as Visual Short-Term Memory. The procedure adopted to display the pictures during the study phase, the Rapid Serial Visual Presentation Procedure (RSVP), seemed particularly suited to study the temporal aspects of recognition memory that are of interest. In RSVP each upcoming picture follows the previous picture in the stream without interruption and serves as an efficient backward mask that affords precise control on the effects of temporal exposure (Potter & Levy, 1969). Pictures can be repeated within the stream with different spacing (by varying the exposure of each single frame) and different temporal lags (1, 2 or 3 intervening pictures) and recognition performance for repeated pictures can be compared with performance for unrepeated pictures within the same stream.

The primary goal of this project was to test for effects of super- and sub-additivity of repetitions against the null hypothesis of total time. Given the apparent differences, methodological as well as of material content, between the present study and classical studies of memory for verbal material, the first question to ask is whether super-additive, facilitatory effects of repetitions, such as spacing and lag effects, will be found in the present visual paradigm. Another possibility is that inhibitory, rather than facilitatory effects may be revealed. For example, in the context of visual short-term memory and RSVP, previous studies have found interference effects between intra-list items, such as Repetition Blindness (Kanwisher, 1987) and

the Attentional Blink (Raymond, Shapiro & Arnell, 1992). Yet, it is an unanswered empirical question whether similar interference effects would appear in the present study, given that the old-new recognition test used here does not require explicit repetition detection or item identification.

We tested recognition memory in two different settings. The first setting was comprised of naïve participants that saw the pictures only once, so that contamination from long-term memory was minimized, but the analysis had to be carried out on pooled data. In the second setting single participants saw the pictures' set repeatedly, thus there was buildup of long-term memory for the seen pictures that contaminated performance, but we could analyze individual subjects' data. The ability to probe differences between single subject and group analysis is important, given the possibility of artifacts induced by averaging across subjects. It is also interesting to test what effect, if any, does long-term memory contamination produce on short-term memory performance.

Previous memory studies have been analyzed and interpreted within two antagonistic theoretical frameworks: Threshold models and Signal Detection Theory (SDT) models. Merits and pitfalls of these two modeling approaches and their variants have long been topics of lively discussions in the memory literature (Wixted, 2007, Yonelinas & Parks, 2007). Here we provide analyses based on both classes of models. It will be evident from the data that both Threshold and Signal Detection Theory accounts provide the same answer in terms of the effects of repetitions on visual short-term recognition memory performance. To anticipate, the results of the present study show no evidence of either superadditive facilitatory effects, such as spacing and lag effects, or inhibitory effects due to intra-list interference: they are instead entirely compatible with the Total Time Hypothesis.

METHODS

Subjects. Two groups of participants were used: 64 undergraduate University students served as naïve participants; an additional 8 undergraduate students took part as practiced participants. All participants were 18-25 years of age and sexes were equally represented in both groups.

Stimuli. Stimuli were 340 colour photographs of generic real life scenes taken at a variety of focal lengths, such as would normally be encountered in everyday life by exploring the visual environment, including outdoor and indoor scenes, objects, people etc. These pictures were presented surrounded by a black background on the face of a CRT monitor with a refresh rate of 75 Hz, in a room with low ambient lighting. At the viewing distance of 57 cm the pictures subtended 8x10 degrees of visual angle.

Procedure: naïve participants. The task consisted of a block of trials, where each trial was comprised of a study and a test phase (figure 1). In the study phase subjects saw an RSVP stream of 9 pictures, comprising an unrepeated image, three repeated images where the repetition followed one, two or three intervening pictures (lag 1, 2, 3, respectively), and two buffer images that were never tested for recognition, one shown at the beginning and one at the end of the stream. After a 1.5 s blank delay following the study phase, a test phase commenced where 8 images were shown singly, the 4 old pictures from the study phase and 4 new images. The observers were given unlimited time to enter their responses on the computer keyboard and indicate whether each picture was new or old. Their reaction time was recorded and they were given no feedback on the accuracy of their responses. Observers also indicated their level of

confidence, by scoring the response (old or new) on a scale 1-3 (low, medium or high confidence, respectively). Each block consisted of 32 trials (preceded by two practice trials), with 4 trials at each of 8 temporal exposures (13, 27, 56, 110, 220, 430, 860 and 1710ms). The framerate of the RSVP stream did not change during a single trial, therefore temporal exposure varied between, but not within trials (between lists factor), whereas repetition and lag varied within each trial (within list factor). Each subject run only one block of trials, thus each subject saw each “new” picture only once (at test), each unrepeated “old” picture twice (once at study and once at test) and each repeated “old” picture three times (twice at study and once at test). The design was counterbalanced over serial study and test position, new and old, repetition, lag and temporal exposure. As such, across all participants each individual image was shown at each temporal exposure as new and old and at each lag.

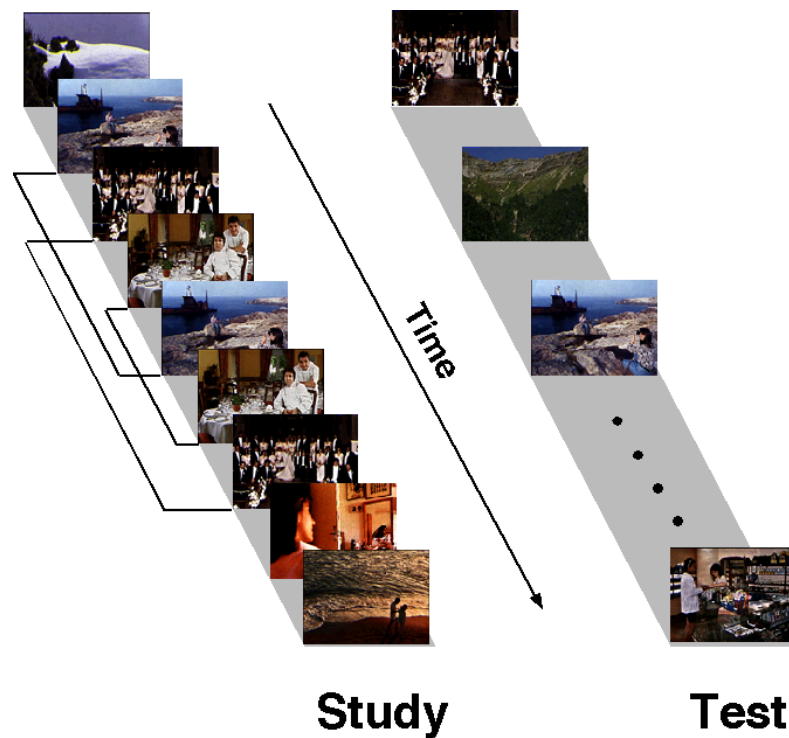


Figure 1. Task procedure for naïve participants. During the study phase subjects saw an uninterrupted stream of 9 images, where each image had the same exposure (chosen from 8 possible exposures, from 12.5msec/picture to 1704msec/picture). Within the stream one image was presented singly and 3 images were repeated with one, two or three intervening images (lag 1, 2, 3, represented by brackets in the figure). Subjects were then tested with 4 images from the stream and 4 new images (the first and last images in the study stream were never tested) and indicated after having seen each image whether it was old or new and how confident they were of their responses on a three-level scale.

Procedure: practiced participants. The design for practiced subjects differed in two crucial aspects from that adopted for naïve participants: only lag 1 repetitions were used and more importantly the pictures' pool was re-sampled several times, so that each participant became highly familiar with the entire set of images. The RSVP stream was comprised of 12 pictures: there were 3 buffer pictures at the beginning and end of the stream that were never tested, one picture repeated with one intervening item (lag 1) and 4 unrepeated pictures. During the test phase subjects were queried on 6 items, 3 "new" pictures and 3 "old" pictures from the RSVP stream, comprised of the repeated picture and 2 of the 4 unrepeated pictures. Reaction time and confidence level were also recorded. As for naïve participants, temporal exposure was a between-trials factor and repetition a within-trial factor. For each trial, the required 14 pictures were drawn at random from the pool of 340 images, such that each picture could be seen more than once by each subject, both within and across blocks, and serial study and test positions were assigned randomly on every trial. Each block comprised 91 trials, with 7 trials at each of 13 exposures (13-3850 ms) and each subject ran 6 blocks over several days.

Data Analysis: Threshold and Signal Detection Theory models.

We provide here a cursory overview of the salient aspects of Threshold and Signal Detection Theory models of memory that are relevant and useful for an immediate understanding

of the present findings. Thorough treatments can be obtained from standard sources (Egan, 1975, Green & Swets, 1966, Macmillan & Creelman, 2005, Wickens, 2002).

Threshold models treat memory as a probabilistic process that allows the observer to be in one of a finite number of states. For example, memory for a test item in High-Threshold models is either above threshold and thus forces the observer to produce an “old” response, or it is below threshold and forces the observer either to declare the item “new” or guess an “old” response. Guessing an “old” response produces one of two possible outcomes: a false alarm or a hit. A false alarm is an incorrect “old” response to a “new” item, whereas a response to an item that has been guessed correctly as “old”, despite the memory signal being below threshold, produces a hit. This contamination of hits by guessed responses unrelated to the memory signal is accounted for in this class of models by subtracting the false alarm rate from the hit rate, thus obtaining an adjusted hit rate that is thought to be related to the strength of the memory signal:

$$\text{Adjusted hit rate} = [\text{hits} / (\text{hits} + \text{misses})] - [\text{false alarms} / (\text{false alarms} + \text{correct rejections})]$$

SDT models of memory are based on the concept of familiarity, a graded quantity related to memory strength that unlike the discrete, probabilistic representation of threshold models is assumed to be a continuous variable. SDT theory assumes that memory signals for “new” and “old” pictures form two Gaussian distributions that differ in location along the familiarity axis, thus having different familiarity means. Unequal Variance SDT models further assume that the “new” and “old” distributions differ in variance, with familiarity of “old” pictures being higher on average, but also more variable than “new” pictures. Threshold, Equal and Unequal-Variance SDT models make different predictions about the distribution of data on a Receiver Operating

Characteristic (ROC) Curve. A ROC curve is constructed by plotting the hit rate versus the false alarm rate at different levels of bias. These coordinate pairs are computed by aggregating responses according to their associated confidence scores. Assuming that the underlying distributions are Gaussian, hit and false alarm rates can then be transformed to z-scores obtaining a z-ROC representation. z-ROC predictions differ among the models: Threshold models predict curvilinear z-ROCs, whereas SDT models predict linear z-ROCs with slopes that are either unity in the Equal Variance case, or less than unity when the variance of “old” is greater than “new” items, as in Unequal Variance models. When z-ROCs are linear (the nearly universal finding with recognition memory data) the intercept represents the difference between the “old” and “new” means divided by the “old” standard deviation $\left(\frac{\mu_o - \mu_n}{\sigma_o}\right)$ and is taken as a measure of familiarity (analogous to the equal variance model d'); the slope is the ratio of “new” to “old” standard deviations $\left(\frac{\sigma_n}{\sigma_o}\right)$, thus a slope less than unity indicates that “old” variance is greater than “new” variance. Dividing the intercept by the slope and taking the inverse of the slope yields, respectively, measures of familiarity and variance in units of “new” standard deviation.

RESULTS

Accuracy: Threshold Theory treatment

For each single exposure, data were pooled across all naïve subjects and the adjusted hit rate was calculated as the difference between hit rate and false alarm rate. An exponential growth model was then fitted by non-linear regression to the obtained scores: $y(t) = m\left(1 - e^{-\frac{t}{\tau}}\right)$.

Exponential growth models have been shown previously to provide excellent fits to recognition memory data for generic visual material in RSVP tasks (Maljkovic & Martini, 2005).

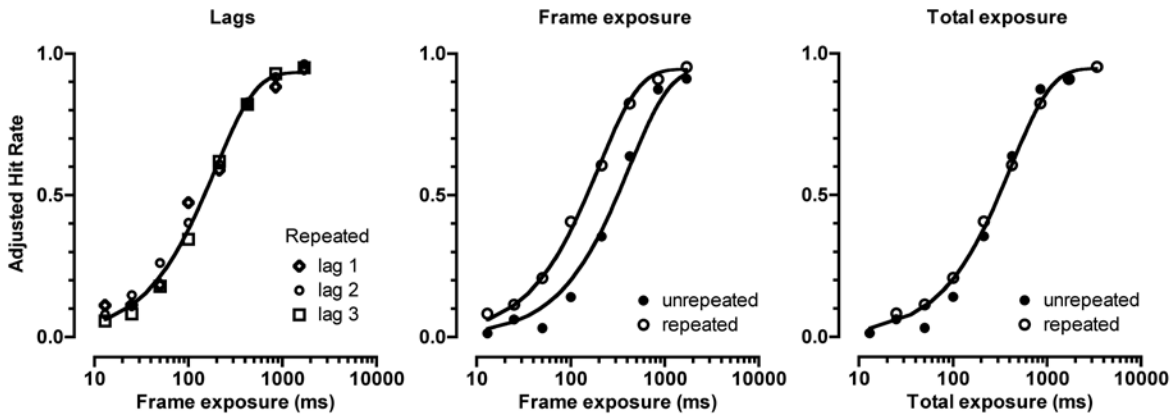


Figure 2. Accuracy data for naïve subjects. **Left**, the adjusted hit rate is plotted as function of frame exposure (duration of a single picture frame in the RSVP stream) for the repeated pictures at all lags. There is no significant difference between lag conditions. The solid line through the data is the exponential model fitted to all 3 conditions with time constant $\tau=192\pm9$ ms, $R^2=0.99$. **Middle**, results for unrepeatd pictures (solid symbols) and repeated pictures (hollow symbols) collapsed across lags, as function of frame exposure. Best fit estimates of the exponential time constants are 418 ± 32 ms and 198 ± 15 ms. **Right**, Results for unrepeatd and repeated pictures are re-plotted as a function of total exposure, demonstrating that in these coordinates data from the two conditions overlap. Best fit time constant is $\tau=411\pm25$ ms, $R^2=0.99$.

Data and model for the repeated conditions with naïve subjects are shown in figure 2, left. Increasing temporal exposure improves recognition memory, leading to almost perfect performance at exposures of about 2 s/picture. However, the rate of improvement is the same for the three lags at all exposures tested: asymptotes and time constants of the fitted exponential models do not differ between lag conditions ($F(2,48)=0.94$, $p=0.46$). Given the apparent absence of a lag effect, data were then collapsed across lags, yielding a single distribution representing memory for repeated pictures. Exponential growth models were then fitted to the unrepeatd and repeated data, as shown in figure 2, middle, obtaining estimates for the exponential time

constants of 418 ± 32 ms and 198 ± 15 ms, respectively. The fact that the time constants are nearly in a 2:1 relationship indicates that memory strength accrued in two separate exposures to the same item is a simple function of total exposure. This is demonstrated more clearly by comparing repeated and unrepeated performance on a common scale of total exposure. As shown in figure 2, right, on such scale repeated and unrepeated performance are indistinguishable and asymptotes and time constants of the fitted models do not differ ($F(2,12)=0.36$, $p=0.7$).

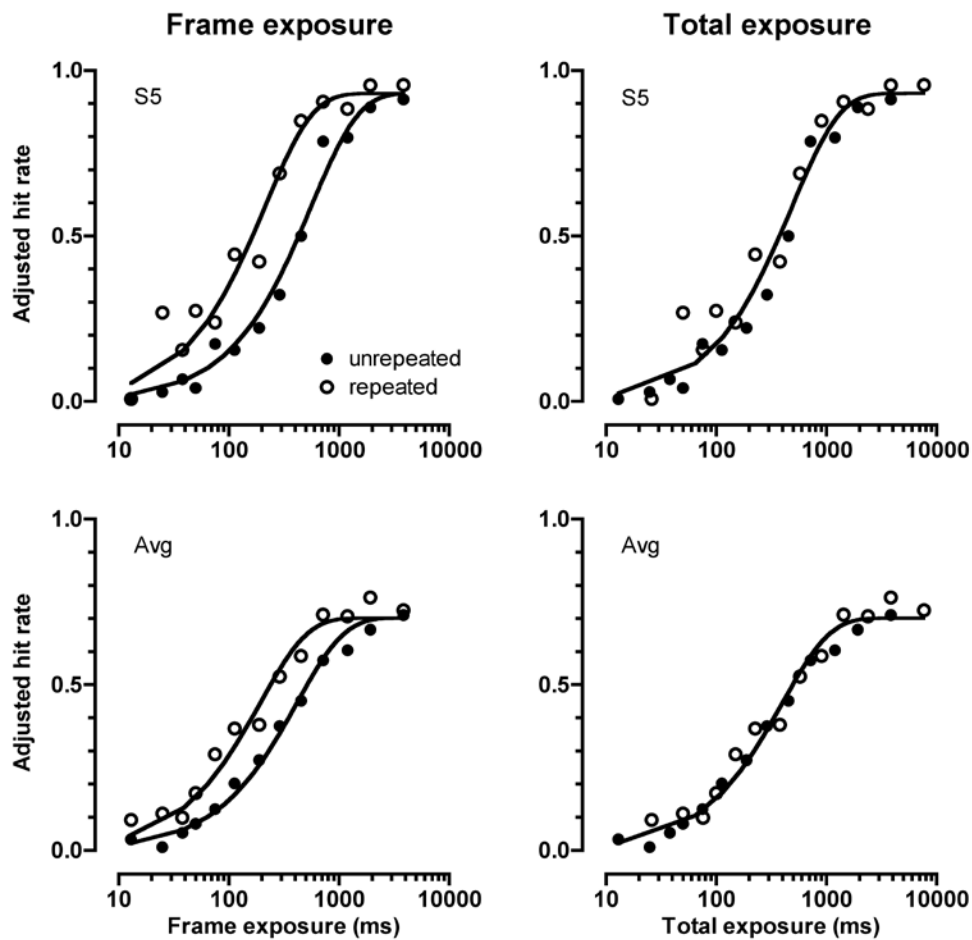


Figure 3. Accuracy data for practiced participants. **Top row**, results are reported for a representative single subject, **bottom row** results are averaged across 8 participants. Single subject's and average results follow similar trends. **Left**, data plotted as function of frame exposure indicate that performance improves twice as fast for repeated than unrepeated pictures, for the single subject as for the average group. **Right**, when re-plotted as function of total exposure data collapse into a single distribution, as for naïve

participants. Notice that asymptotic performance is on average lower for practiced than naïve participants, but the time constants are similar.

Very similar results are obtained with practiced participants, as shown in figure 3, top, for a representative single subject, and in figure 3, bottom, as the aggregate performance of all 8 participants. As with naïve participants, the time constants of the exponential model fitted to unrepeated and repeated performance scores are nearly in a 2:1 relationship (single subject: 558 ± 58 ms and 213 ± 22 ms, average of subjects: 413 ± 33 ms and 192 ± 15 ms, see figure 3, left) and data from the two conditions overlap on a common scale of total exposure (shared time constant 398 ± 25 , figure 3, right). Although the time constants are very similar to those of naïve participants, practiced subjects achieve on average a lower asymptotic performance at long exposures (compare figure 2, right, with figure 3, bottom right). This may relate to the main difference between the two designs: while naïve participants saw each picture only once, twice or three times (“new”, “old” unrepeated and “old” repeated, respectively), for practiced participants each picture in the pool had the same probability of being chosen for display on every trial and thus could be seen several times across trials within a single block and across blocks, both as “new” and “old”. Therefore, in the practiced participants’ case, differences in familiarity between “new” and “old” pictures were superimposed on a higher baseline of long-term memory elicited by previous encounters. We return to this point below.

Accuracy: Signal Detection Theory treatment

SDT analysis was carried out on the data pooled across participants. Following standard procedures (Wickens, 2002), z-ROC curves were constructed from z-transforms of hit and false alarm rates across different levels of confidence, separately for each exposure (see figure 4).

Straight lines were fitted to the data by linear regression, taking into account variability in both x and y dimensions. There was no significant deviation from linearity, suggesting normality of the underlying distributions.

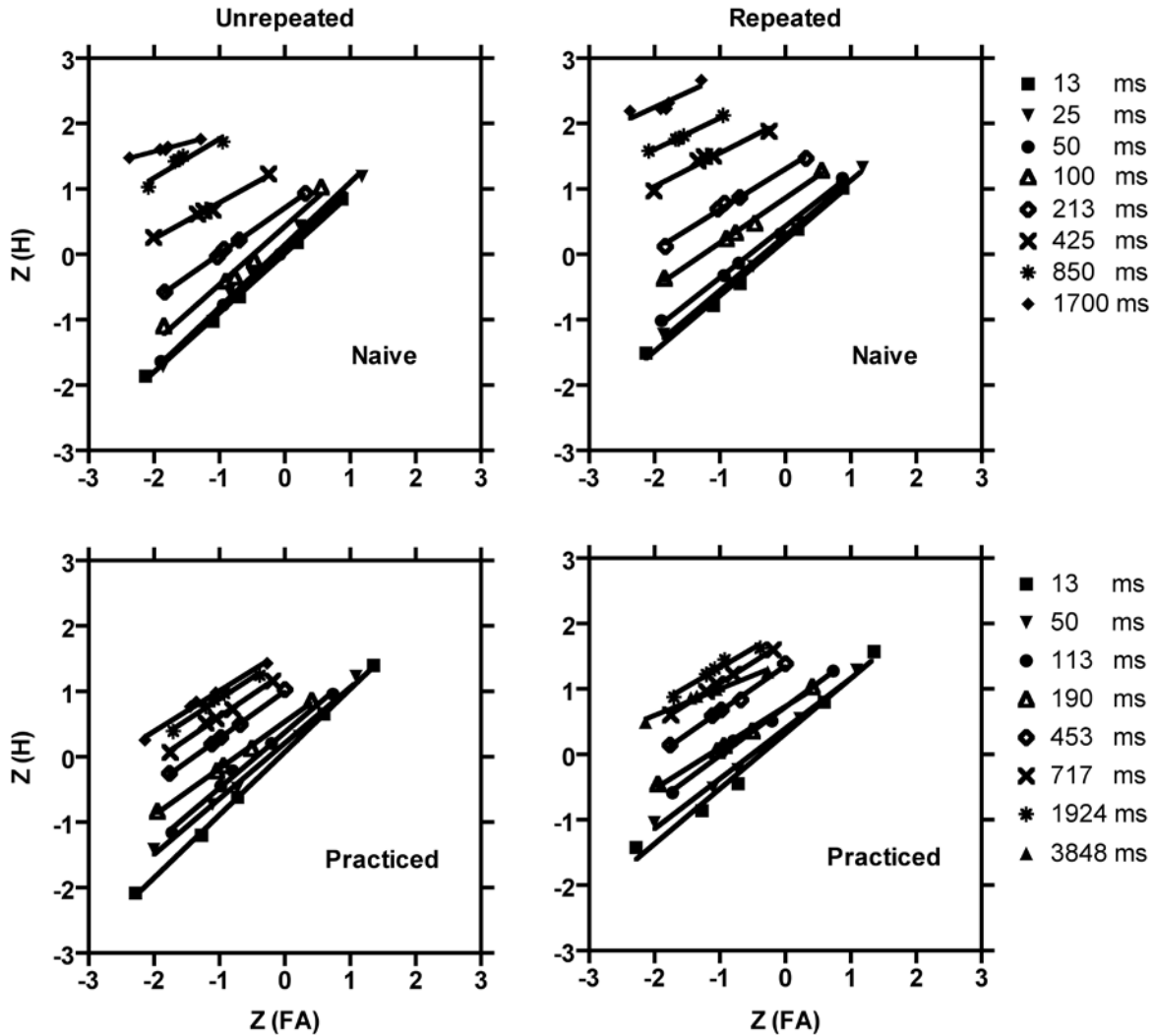


Figure 4. **Left**, zROC curves for unreported and **right**, repeated pictures; **top**, naïve and **bottom**, practiced participants. In each diagram, the z-scored hit rate is plotted as function of the z-scored false alarm rate corresponding to the 5 criteria separating responses with different levels of confidence. Straight lines are linear regressions applied to the data, accounting for variance in both x and y, for each exposure. No significant deviations from linearity were observed, suggesting normality of the underlying distributions. Consistent with improving performance, the curves' distance from the diagonal and their intercepts increase with increasing exposure. Conversely, the curves' slopes decrease with exposure.

Figure 5 shows intercepts and slopes of the fitted z-ROC curves as function of temporal exposure, for repeated and unrepeated pictures and for naïve and practiced participants. Intercepts, which represent a measure of familiarity in units of “old” standard deviation, increase with exposure and are higher for repeated than unrepeated pictures (figure 5, top). Slopes, which represent the ratio of “new” to “old” standard deviations, are all less than unity, they are smaller for repeated than unrepeated pictures and decrease with exposure (figure 5, bottom). These results are qualitatively similar for both groups of participants.

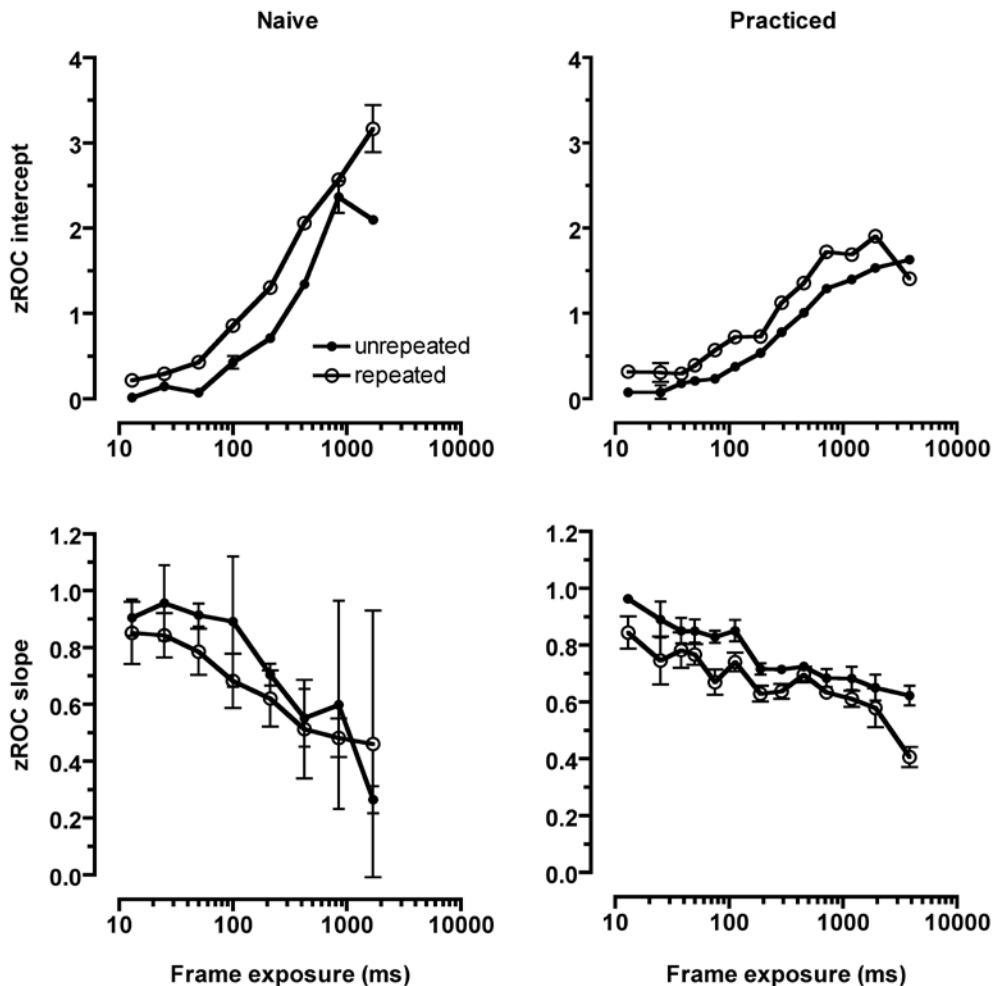


Figure 5. **Top row**, z-ROC intercepts and **bottom row**, slopes are plotted for unrepeated (solid symbols) and repeated pictures (hollow symbols) as function of frame exposure; **left**, naïve and **right**, practiced

participants. The scale for both intercepts and slopes is in units of “old” standard deviation. Errors are 95% C.I. Intercepts increase monotonically with exposure and are higher for repeated than unrepeated pictures. Inversely, slopes decrease monotonically with exposure and are lower for repeated than unrepeated pictures. The results imply that average “old” familiarity and its standard deviation increase with exposure and are higher for repeated than unrepeated pictures. Data for naïve and practiced participants show the same trends.

Assuming that the variance of the distribution of familiarity of “new” pictures does not change systematically with exposure (a reasonable assumption given the mixing of exposures and repetition condition in the experimental design), this indicates that the standard deviation of the familiarity of “old” pictures increases with exposure and that the familiarity of repeated pictures is more variable than unrepeated pictures. Dividing the intercepts by the slopes, thus expressing familiarity in common units of “new” standard deviation, and plotting these scores as a function of total exposure, reveals that data for repeated and unrepeated pictures overlap, indicating that total exposure fully predicts the familiarity of repeated pictures (figure 6). While the curves for naïve (figure 6, left) and practiced participants (figure 6, right) have similar shapes, the growth of the data for practiced subjects is attenuated and at long exposures performance remains lower than naïve subjects by about a factor-of-two. Perhaps the simplest explanation of this finding is that the standard deviation of the familiarity of “new” pictures, which determines the scale of the ordinate, differs between the two paradigms, being greater for practiced than naïve participants due to the repeated sampling of the pictures’ pool across trials. As such, the higher variance of “new” pictures scales down the performance of practiced compared to naïve participants.

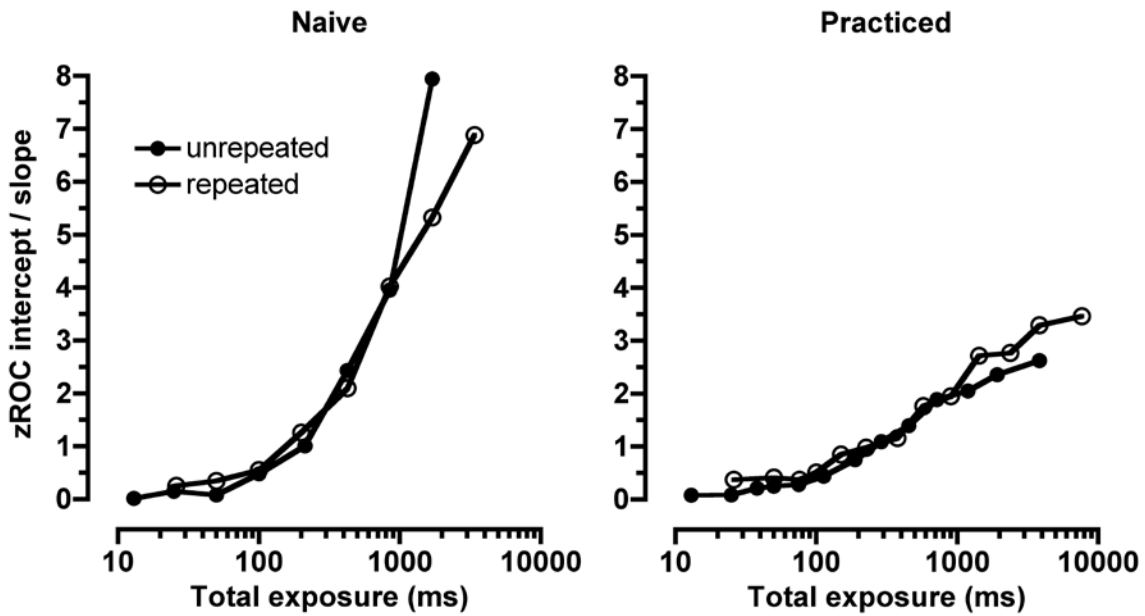


Figure 6. **Left**, accuracy data for naïve and **right**, practiced participants, with repeated (solid symbols) and unrepeated pictures (hollow symbols). Memory performance, expressed as distance of the “old” from the “new” distribution in units of “new” standard deviation (z-ROC intercepts divided by slopes), is plotted as function of total exposure to the pictures. The two distributions overlap, indicating that total exposure completely determines performance in the repeated condition. The distribution for practiced participants appears as a scaled down version of naïve participants, suggesting differences in the scale of the ordinate. This is consistent with the hypothesis that the standard deviation of “new” pictures is higher for practiced than naïve subjects, due to differences in the experimental design.

Analysis of the z-ROC curves of naïve participants for repeated pictures at different lags reveals no effect of lag on either intercepts or slopes (figure 7).

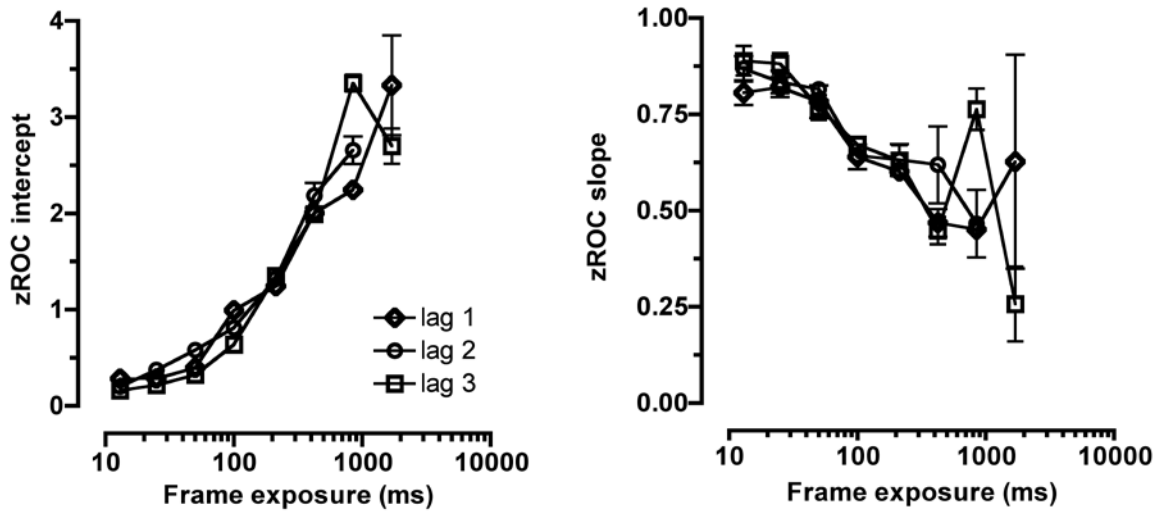


Figure 7. **Left**, z-ROC intercepts and **right**, slopes are plotted as function of exposure to repeated pictures for naïve participants in the 3 lag conditions. Errors are 95% C.I. Neither intercepts nor slopes differ between lags.

Finally, in figure 8 the inverse of the slopes are plotted as function of the intercepts divided by slopes, for all data sets including unrepeated and repeated pictures and naïve as well as practiced participants. This diagram represents the relationship between standard deviation and mean of the familiarity distribution of “old” pictures in “new” standard deviation units. A power law model $y=1 + \alpha x^\beta$ was fitted to the data, obtaining estimates $\alpha=0.39$ (C.I. 0.34-0.44) and $\beta=0.59$ (C.I. 0.48-0.70). Separate analysis of the four individual data sets (repeated and unrepeated for naïve and practiced participants) yields statistically indistinguishable estimates of the model’s parameters. The power law exponent is not significantly different than 0.5, indicating that the *variance* of the familiarity distribution of “old” pictures grows linearly with the mean. This may be taken to indicate that the noise associated with “old” items is uncorrelated (or very little correlated).

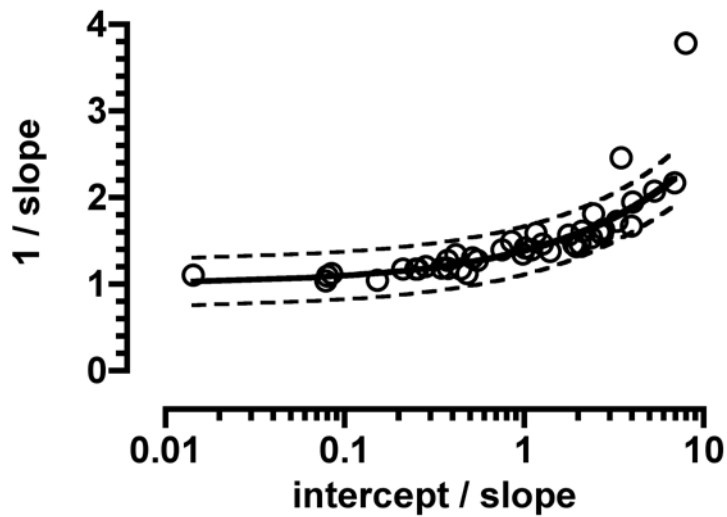


Figure 8. The inverse of the z-ROC slope is plotted as function of the intercept divided by the slope. Data are for unrepeated and repeated pictures, from both naïve and practiced participants. The scale of both axes is in units of “new” standard deviation. The diagram represents the growth of the standard deviation with the mean of the familiarity distribution of “old” pictures. The solid line is the best fit of a power function to the data; the dashed lines indicate the 95% prediction interval. The power exponent is not significantly different than 0.5 (the *variance* of the distribution grows linearly with the mean).

In summary, the evidence for both naïve and practiced participants and for single subjects as for aggregate data, indicates that visual short-term recognition memory for generic visual scenes in the RSVP paradigm is consistent with the Total Time Hypothesis and lacks superadditive facilitatory effects of spacing and lag, as well as inhibitory effects normally found in repetition blindness and attentional blink paradigms.

Reaction times and confidence ratings

Reaction times and confidence ratings for naïve participants are shown in figure 9 (similar results, not shown, were obtained with practiced participants). For each total exposure, data were pooled across participants and repetition conditions.

Correct “old” responses (Hits) become progressively faster with increasing temporal exposure to the pictures. An exponential decay model $y = me^{-\frac{t}{\tau}} + c$ with a fixed decay parameter $\tau=411$ ms, as obtained from the adjusted hit rate data (as in figure 2, right), provides an excellent fit to Hit response times (figure 9, top, left). False positive responses (FA, false alarms) are as slow as correct responses to pictures seen very briefly. Average confidence ratings for Hit responses increase with exposure and for False Alarms they are as low as Hits to very briefly seen pictures (figure 9, left, bottom). Taken together, these results suggest that familiarity, speed and confidence are very highly correlated in the case of “old” responses.

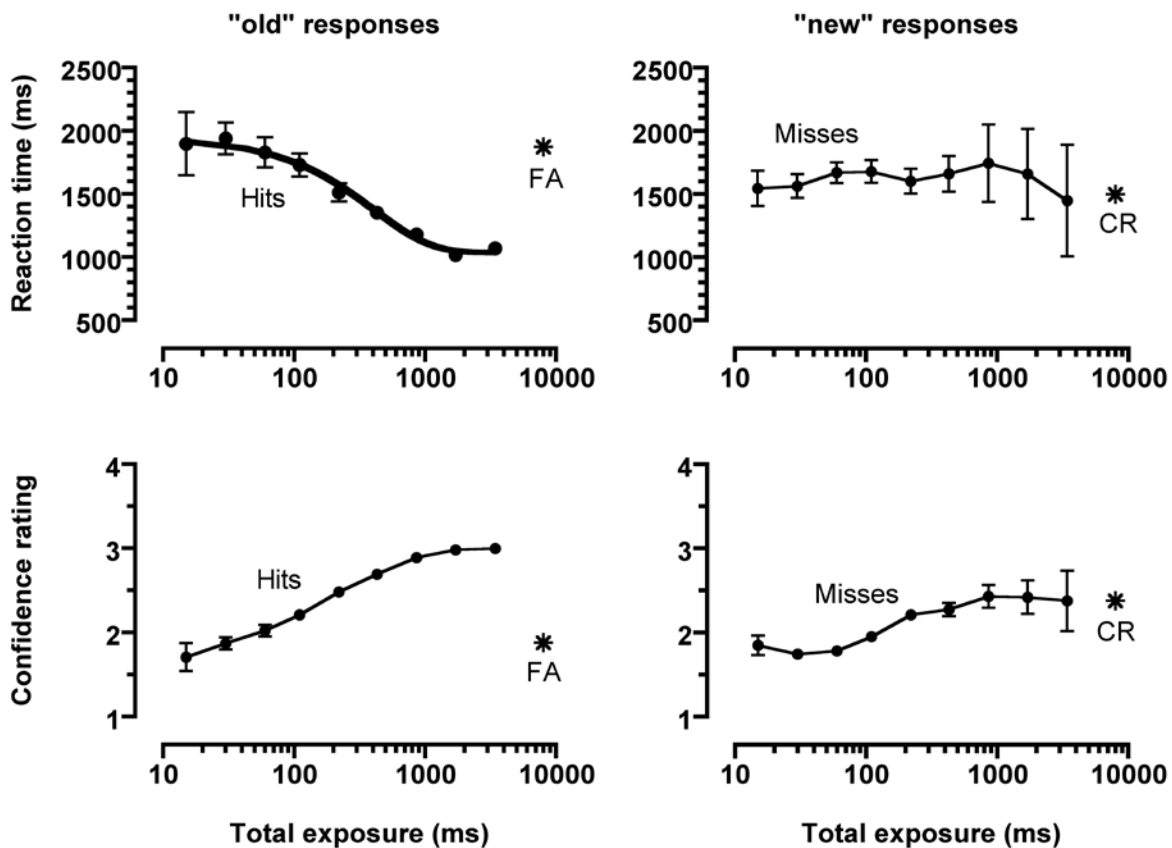


Figure 9. **Top**, reaction times and **bottom**, average confidence scores of naïve participants are plotted as a function of total exposure, collapsed across repeated and unrepeated pictures, **left** for “old” and **right**, for “new” responses. FA are false alarms; CR are correct rejections. Errors are 95% C.I. The solid line

through the reaction times for hits (top, left) is an exponential decay function with the same time constant as the accuracy data in figure 2, right ($\tau=411$ ms). There is an almost perfect correlation between speed, accuracy and confidence for “old” responses. Reaction times for incorrect “new” responses to “old” pictures (misses, left, top) do not depend on exposure, but confidence scores increase with exposure (left, bottom). The increase in variance with familiarity and shifts in criteria explain the temporal dependence of confidence scores for errors.

Differently than “old” responses, the speed of incorrect “new” responses to “old” pictures (Misses) does not depend on temporal exposure and is similar to the speed of correct “new” responses (CR, correct rejections) (figure 9, right, top). However, incorrect responses to “old” pictures seen briefly are committed with less confidence than to pictures seen at long exposures, where the reported confidence is the same as observed with correct rejections (figure 9, right, bottom). In the context of a SDT account, two factors may explain this counterintuitive growth of confidence in errors with temporal exposure. The first factor to consider is the “old” distribution’s variance. As exposure increases, the distribution of familiarity of “old” pictures shifts away from the distribution of “new” pictures improving the hit-rate, but at the same time the distribution also grows in variance, increasing the relative probability of encountering “old” pictures with extremely low familiarity that are incorrectly classified as “new”. The progressive fattening of the left-hand tail of the “old” distribution increases the relative proportion of “old” samples falling below the high confidence “new” criterion. Secondly, confidence criteria may change with exposure. The location of the criteria on the familiarity axis can be computed by negating the z-scores of the false alarm rates associated with different levels of confidence (by negating the x coordinates of the z-ROC curves) (Stretch & Wixted, 1998a, Wixted & Gaitan, 2002). As such, criteria are expressed as distances from the mean of the “new” distribution, in units of “new” standard deviation. The criteria used by naïve and practiced participants at different temporal exposures are reported in figure 10, top. For both participants’ groups, there is

a tendency for the distance between criteria to narrow as exposure increases and this tendency is more pronounced for the leftmost criteria (NH and NM, associated with “new” responses) than for the rightmost criteria (OH and OM, associated with “old” responses), resulting in an asymmetrical shift towards higher familiarity levels. This shift means that with growing exposure observers become more conservative, reducing the false alarm rate, but also increasing the probability of missing “old” pictures with high confidence.

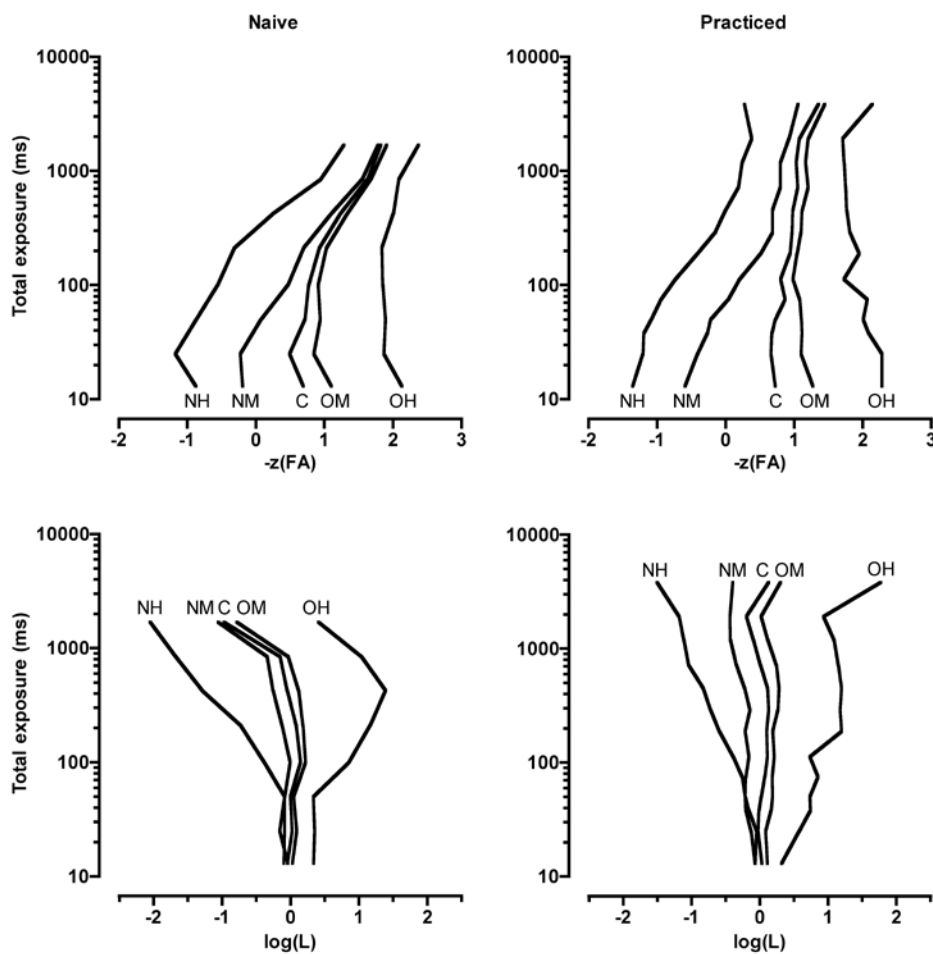


Figure 10. **Top**, the location of criteria on the familiarity axis, expressed as distance from the mean of the “new” distribution in units of “new” standard deviation, is plotted for each temporal exposure studied, **left** for naïve and **right** for practiced subjects. Labels for criteria are N for “new” and O for “old”; H for high confidence and M for medium confidence; C is the central criterion separating “new” from “old” responses. The observed “fanning” pattern could be qualitatively consistent with an effort to maintain a constant likelihood ratio of a picture being “old” to “new”. This conjecture is tested in the **bottom** graphs of the figure. The log of the likelihood ratio corresponding to the observed criteria is plotted for all

exposures studied. The central criterion is maintained at a constant likelihood ratio (equal likelihoods of “old” and “new”), but the more extreme criteria fan out with increasing exposure.

One interesting question regards the mechanism by which observers choose the locations of criteria. An influential account of criterion placement, going back to the origins of SDT (Green & Swets, 1966), is that criteria represent likelihood ratios of pictures being “old” versus “new”. For example, the optimal placement of the criterion separating “old” from “new” items corresponds to the point along the familiarity axis where the probability of a picture being “old” is equal to the probability of being “new”, corresponding to a likelihood ratio of unity. If observers choose their criteria based on fixed likelihood ratios, then as memory strength increases with exposure criteria should move on the familiarity axis in a manner qualitatively consistent with the “fanning” pattern observed in the data reported in figure 10, top. To see if the observed pattern is indeed quantitatively consistent with this invariance hypothesis, likelihood ratios corresponding to the reported confidence criteria were computed for each exposure (by using equation A4 in (Stretch & Wixted, 1998a)). Figure 10, bottom, reports how the logarithm of the computed ratios ($\log(L)$) changes with the logarithm of temporal exposure. In this coordinate space, the log-likelihood of the central criterion C separating “old” from “new” pictures remains roughly invariant at 0 with growing exposure, consistent with the criterion being placed optimally. The more extreme criteria for “new” and “old” pictures, however, tend to increase their distance with exposure in a roughly linear (in this double logarithmic representation) and symmetrical manner, implying more conservative judgments at longer exposures. If the underlying assumptions of the analysis are correct (particularly if the standard deviation of the “new” distribution does not change with exposure), then clearly this analysis

shows that observers do not maintain their most extreme criteria at fixed likelihood ratios as memory strength increases.

Discussion

Absence of interference and distributed practice effects in RSVP recognition tasks

The literature concerning studies that have used the RSVP paradigm for stimulus presentation has been dominated in the past several years by two influential findings and associated theoretical accounts: Repetition Blindness (RB) and the Attentional Blink (AB). Both these phenomena represent interference effects between intra-list items. RB refers to the finding that observers fail to recognize that an item has been repeated within a list when the repetition occurs with a short delay of about 150 ms (Kanwisher, 1987); it is a robust effect with verbal material and has been reported also with streams of pictures (Bavelier, 1994, Coltheart, 1999), although the effect with pictures seems less reliable. AB refers to the finding that a second target is frequently missed when it appears 200-500 ms after the onset of a first target and it is a robust effect with pictures (Einhauser, Koch & Makeig, 2007, Raymond et al., 1992). While interference effects have been demonstrated with RSVP sequences similar to those used in the present study, it is important to recognize the differences in task demands: the present recognition task required only an old/new judgment, whereas RB and AB tasks require identification and explicit report of the targets. Repetition of a picture in the present study always led to improved recognition performance, of a magnitude such as to exclude interference between items. In one of the earliest RSVP studies (Potter & Levy, 1969) Potter and Levy made a related observation, based on sequential analysis: they noted that recognition of a picture did not interact with recognition of the immediately following picture in the RSVP stream. It seems

plausible to assume that the absence of interference effects in recognition as opposed to identification paradigms must be related to differences in task demands. Future investigations inspired by such differences may contribute to clarify the nature of the interferences found in RB and AB tasks, which at present remain controversial. From a pragmatic point of view, the present results simplify the discussion of short-term recognition memory mechanisms by demonstrating that intra-list interference effects are not detectable and therefore unlikely to play a role in judgments of picture familiarity.

A second topic that has been prominent in the recent memory literature regards the facilitatory effect of distributed practice. Increasing lag (number of intervening items) and spacing (temporal distance between repetitions) has been shown to improve the effectiveness of learning beyond the level produced by an equivalent massed study time (Kahana & Howard, 2005), but the majority of relevant studies have been concerned with recall tasks of verbal material and relatively long retention intervals (Cepeda et al., 2006). In the present study of recognition memory with pictures there were no effects of either lag or spacing: recognition performance with repeated pictures did not vary with the number of intervening items and was indistinguishable from performance obtained with an equivalent massed study time. The same result was borne out by measuring adjusted hit rates, based on a Threshold model analysis, as by a more sophisticated measure of familiarity, based on an Unequal Variance SDT analysis. We do not know of any study that has used words in the same repetition paradigm used here. As such it remains unclear whether the absence of distributed practice effects under the present conditions is specific for pictures or generalizes to other material.

Total Time Hypothesis and memory mechanisms

The obtained results are consistent with the Total Time Hypothesis (Bugelski, 1962), which states that “a fixed amount of time is necessary to learn a fixed amount of material regardless of the number of individual trials into which that time is divided” (Cooper & Pantle, 1967). As such, the hypothesis is simply a descriptive device: what do the results tell us about memory mechanisms?

Consider the implications in terms of memory decay. The absence of lag or spacing effects and the total time dependence suggest that no appreciable memory decay exists over a span of several seconds and several intervening items. Potter and colleagues (Potter, Staub, Rado & O'Connor, 2002) have shown that there are no appreciable serial *study* position effects with RSVP streams of up to 10 items at 5.5 Hz framerate, suggesting lack of decay over 10 intervening items and/or at least 2 s. Notice that in the same study robust serial *test* position effects were observed, consistent with previous findings with verbal material and longer retention intervals (Kim & Glanzer, 1995), a pattern that may suggest an explanation of memory decay at test in terms of retrieval interference (Anderson, Bjork & Bjork, 1994, Roediger, 1974). Melcher (Melcher, 2001, Melcher, 2006) observed total time dependencies of recognition memory over a span of up to 20 seconds and several intervening items and distracting tasks, but also noted that no memory buildup could be found across days. Older (Shepard, 1967, Standing, 1973, Standing, Conezio & Haber, 1970) and more recent studies indicate an impressive capacity for picture memory: in the most recent study by Brady and colleagues (Brady, Konkle, Alvarez & Oliva, 2008) 2500 pictures were shown at a rate of 3 s/picture and recognition memory remained above 80% at lags of 1000 intervening pictures. In addition, the same study showed that the information retained is sufficiently fine grained to allow discrimination of detail, not just the gist of the image. The available evidence thus suggests that we are concerned here with a

memory system capable of supporting recognition of hundreds (perhaps thousands) of complex images over a span of seconds (perhaps minutes) and/or hundreds (perhaps thousands) of intervening items.

Encoding capacity and decay rate are tightly coupled parameters in memory networks (Amit & Fusi, 1994, Yakovlev, Amit, Romani & Hochstein, 2008). Evidence from Neuroscience suggests that memories entail modifications of the strength of bounded synapses and that memory decay is due to overwriting of these synaptic modifications by ongoing plasticity, either spontaneous or due to interference by other memories, in a spike-timing dependent manner. The number of synapses in the network and the degree to which activity evoked by a stimulus overlaps with activity evoked by previous stimulus encounters are two factors that influence both the capacity and the decay rate (Fusi & Abbott, 2007). The behavioral results with picture memory, consistent with high capacity and relatively slow decay, suggest that pictures must be represented sparsely in memory, such that interference is kept to a minimum. While much recent research indicates that a major goal of early vision is to reduce redundancy (Simoncelli & Olshausen, 2001) and achieve sparse coding to optimize metabolic energy efficiency (Attwell & Laughlin, 2001, Levy & Baxter, 1996, Olshausen & Field, 2004), it is clear that one major advantage of decorrelated and sparse image codes is to allow optimal memory encoding (Willshaw, Buneman & Longuet-Higgins, 1969).

Accumulation of memory strength as a stochastically independent process

One of the most robust findings of this as well as other studies on the growth of short-term memory performance with temporal exposure concerns the form of the best fitting function describing such growth. Fraction correct or adjusted hit rate grows with exposure exponentially

(Lamberts, Brockdorff & Heit, 2002, Maljkovic & Martini, 2005). This suggests a notion of independence: what is added to memory in the current instant does not interact with what was added to memory previously beyond what is expected from random coincidences, paradoxically a “memory without memory”. Accordingly, performance in the repeated condition could be accurately predicted from performance in the unrepeated condition via a probability summation formula. Disputing such threshold analysis on account of its well-known inadequacies does not change the conclusion: the same outcome is borne out by the SDT analysis of the growth of noise with familiarity as follows. It is now recognized that the vast majority of recognition memory experiments has produced zROC curves with slopes that are less than unity (Glanzer, Kim, Hilford & Adams, 1999, Heathcote, 2003, Hirshman & Hostetter, 2000, Yonelinas & Parks, 2007), suggesting that the familiarity representation of “old” items has a standard deviation greater than “new” items. Familiarity grows with temporal exposure and so does its standard deviation: as the intercept of the zROC curve increases with exposure, its slope diminishes. This finding raises two questions: by how much does variability increase with familiarity and why does it increase at all? The results of this study (figure 8) show that the standard deviation of the distribution of familiarity of “old” pictures (expressed in units of “new” standard deviation) grows roughly with the square root of the mean, suggesting proportionality of the *variance* with the mean and thus indicating that the noise is uncorrelated. As such, the growth of familiarity with temporal exposure is a process akin to a random walk with drift. In the absence of sufficient mechanistic knowledge of the underlying neurobiology, we can only speculate on the instantiation of such walk at a process level. Possibilities include the progressive recruitment and/or potentiation/depression of a large number of independent elementary mechanisms encoding the stimulus. While on average such growth is positive, leading to an

increased memory signal, the crucial observation is that growth is noisy, resulting sometimes in a strong signal and other times in a signal so weak as to be indistinguishable from baseline noise.

Decision criteria shift with memory strength

There are no discrepancies in the results discussed so far between the Threshold Model and SDT analysis. As such, the reader may be led to conclude with S.S. Stevens that in computing the somewhat more elaborate measures of SDT we have engaged in “much honing of the tool’s edge, but little cutting” (Stevens, 1975). Stevens was a proponent of an extreme version of threshold theory known as the Neural Quantum and in his characteristic style described SDT as follows: “As I understand it, the idea is simply that, when a human observer undertakes to detect a signal immersed in noise, his behavior has much in common with statistical decision theory. Being confronted with two statistical distributions, that of the noise and that of the noise plus the signal, the observer seems to behave as though he were testing a statistical hypothesis: given a sample, he tries to decide which of the two populations it came from. Depending on the ‘pay-off matrix’, the observer may be timid or bold in his willingness to commit errors of one kind or another, and the degree of his daring helps to determine a boundary criterion (a cut-off) for the categories of his response. Since the parameters of the experiment can move the cut-off up or down the ‘decision axis’, there is said to be no unique threshold in the sense of an all-or-none process” (Stevens, 1961). Perhaps against the intention of its skeptical Author, this succinct description of SDT makes clear its distinctive advantage over threshold models: the natural ability to account for processes of decision-making and motivation, as well as for those of perception and memory. In the present study it was found that familiarity, speed and confidence are almost perfectly correlated in the case of “old” responses, a result that is

perhaps as reassuring as it is unsurprising and easily understandable from a threshold model perspective. It is not unreasonable to suppose that a higher familiarity signal increases the probability of a correct response, reduces the time necessary for reaching a threshold that triggers a response and increases confidence inasmuch as confidence is proportional to the familiarity signal. What seems more difficult to explain without resorting to SDT is the finding that confidence grows with exposure also in the case of incorrect “new” responses. A related observation is the so-called “mirror effect”, the finding that higher hit rates are always associated with lower false alarm rates (Glanzer & Adams, 1985). In both cases the explanation has to do with shifts in criteria: as familiarity increases, observers tend to become more conservative (Stretch & Wixted, 1998b). The “fanning” pattern of shifts in criteria that we have observed with increasing exposure (figure 10) is, to a first approximation, qualitatively consistent with an effort to maintain constant likelihood ratios of “old” versus “new” as the familiarity of “old” pictures increases (Stretch & Wixted, 1998a). Maintaining fixed likelihood ratios is what an ideal observer would do, having perfect knowledge at any instant of time of the shape of the distributions, their location and variance (Green & Swets, 1966). Implausible as this may be in the case of the human observer, other mechanisms, shaped by learning, previous history of reinforcement and trial-by-trial adjustments, may mimic this invariance (Treisman & Williams, 1984, Wixted & Gaitan, 2002). Indeed, the central criterion separating “old” from “new” responses seems to shift in a way such as to maintain a constant ratio across exposures, but the more extreme criteria do not. We find that the criteria other than the central one seem roughly to obey a rule such that their distance increases linearly and symmetrically on a log-likelihood ratio scale as function of the logarithm of temporal exposure (see figure 10, bottom). The nature of the underlying computation and the ability of this observation to generalize to other settings (for

example low prevalence search tasks (Wolfe, Horowitz, Van Wert, Kenner, Place & Kibbi, 2007)) remain topics for future investigations.

References

- Amit, D.J., & Fusi, S. (1994). Learning in neural networks with material synapses. *Neural Computation*, 6(5), 957-982.
- Anderson, M.C., Bjork, R.A., & Bjork, E.L. (1994). Remembering can cause forgetting: Retrieval dynamics in long-term memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 20(5), 1063-1087.
- Attwell, D., & Laughlin, S.B. (2001). An energy budget for signaling in the grey matter of the brain. *J Cereb Blood Flow Metab*, 21(10), 1133-1145.
- Bavelier, D. (1994). Repetition blindness between visually different items: The case of pictures and words. *Cognition*, 51(3), 199-236.
- Brady, T.F., Konkle, T., Alvarez, G.A., & Oliva, A. (2008). Visual long-term memory has a massive storage capacity for object details. *Proc Natl Acad Sci U S A*, 105(38), 14325-14329.
- Bugelski, B.R. (1962). Presentation time, total time, and mediation in paired associate learning. *Journal of Experimental Psychology*, 63(4), 409-412.
- Cepeda, N.J., Pashler, H., Vul, E., Wixted, J.T., & Rohrer, D. (2006). Distributed practice in verbal recall tasks: A review and quantitative synthesis. *Psychological Bulletin*, 132(3), 354-380.
- Coltheart, V. (1999). *Fleeting memories: Cognition of brief visual stimuli*. MIT Press/Bradford Books series in cognitive psychology. (p. 271 pp). Cambridge, MA: The MIT Press.
- Cooper, E.H., & Pantle, A.J. (1967). The Total-Time Hypothesis in Verbal Learning. *Psychological Bulletin*, 68(4), 221-234.
- Ebbinghaus, H. (1964). *Memory: a contribution to experimental psychology*. (pp. xvi, 123 p.). New York,: Dover Publications.
- Egan, J.P. (1975). *Signal detection theory and ROC-analysis*. (pp. xx, 277 p.). New York: Academic Press.
- Einhauser, W., Koch, C., & Makeig, S. (2007). The duration of the attentional blink in natural scenes depends on stimulus category. *Vision Research*, 47(5), 597-607.

- Fusi, S., & Abbott, L.F. (2007). Limits on the memory storage capacity of bounded synapses. *Nature Neuroscience*, *10*(4), 485-493.
- Glanzer, M., & Adams, J.K. (1985). The mirror effect in recognition memory. *Memory & Cognition*, *13*(1), 8-20.
- Glanzer, M., Kim, K., Hilford, A., & Adams, J.K. (1999). Slope of the receiver-operating characteristic in recognition memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *25*(2), 500-513.
- Green, D.M., & Swets, J.A. (1966). Signal detection theory and psychophysics. (pp. xi, 455 p.). New York,: Wiley.
- Harris, C.M., Hainline, L., Abramov, I., Lemerise, E., & et al. (1988). The distribution of fixation durations in infants and naive adults. *Vision Research*, *28*(3), 419-432.
- Heathcote, A. (2003). Item Recognition Memory and the Receiver Operating Characteristic. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *29*(6), 1210-1230.
- Hirshman, E., & Hostetter, M. (2000). Using ROC curves to test models of recognition memory: The relationship between presentation duration and slope. *Memory & Cognition*, *28*(2), 161-166.
- Kahana, M.J., & Howard, M.W. (2005). Spacing and lag effects in free recall of pure lists. *Psychon Bull Rev*, *12*(1), 159-164.
- Kanwisher, N.G. (1987). Repetition blindness: Type recognition without token individuation. *Cognition*, *27*(2), 117-143.
- Kim, K., & Glanzer, M. (1995). Intralist interference in recognition memory. *J Exp Psychol Learn Mem Cogn*, *21*(5), 1096-1107.
- Lamberts, K., Brockdorff, N., & Heit, E. (2002). Perceptual processes in matching and recognition of complex pictures. *Journal of Experimental Psychology: Human Perception & Performance*, *28*(5), 1176-1191.
- Levy, W.B., & Baxter, R.A. (1996). Energy efficient neural codes. *Neural Comput*, *8*(3), 531-543.
- Macmillan, N.A., & Creelman, C.D. (2005). Detection theory : a user's guide. (pp. xix, 492 p.). Mahwah, N.J.: Lawrence Erlbaum Associates.
- Maljkovic, V., & Martini, P. (2005). Short-term memory for scenes with affective content. *J Vis*, *5*(3), 215-229.
- Melcher, D. (2001). Persistence of visual memory for scenes. *Nature*, *412*(6845), 401.

- Melcher, D. (2006). Accumulation and persistence of memory for natural scenes. *Journal of Vision*, 6(1), 8-17.
- Olshausen, B.A., & Field, D.J. (2004). Sparse coding of sensory inputs. *Current Opinion in Neurobiology*, 14(4), 481-487.
- Potter, M.C., & Levy, E.I. (1969). Recognition memory for a rapid sequence of pictures. *Journal of Experimental Psychology*, 81(1), 10-15.
- Potter, M.C., Staub, A., Rado, J., & O'Connor, D.H. (2002). Recognition memory for briefly presented pictures: The time course of rapid forgetting. *Journal of Experimental Psychology: Human Perception and Performance*, 28(5), 1163-1175.
- Raymond, J.E., Shapiro, K.L., & Arnell, K.M. (1992). Temporary suppression of visual processing in an RSVP task: An attentional blink? *Journal of Experimental Psychology: Human Perception and Performance*, 18(3), 849-860.
- Roediger, H.L. (1974). Inhibiting effects of recall. *Memory & Cognition*, 2(2), 261-269.
- Roediger, H.L., III (2008). Relativity of remembering: Why the laws of memory vanished. *Annual Review of Psychology*, 59, 225-254.
- Shepard, R.N. (1967). Recognition Memory for Words, Sentences, and Pictures. *Journal of Verbal Learning & amp*, 6(1), 156-163.
- Simoncelli, E.P., & Olshausen, B.A. (2001). Natural image statistics and neural representation. *Annual Review of Neuroscience*, 24, 1193-1216.
- Standing, L. (1973). Learning 10,000 pictures. *The Quarterly Journal of Experimental Psychology*, 25(2), 207-222.
- Standing, L., Conezio, J., & Haber, R.N. (1970). Perception and memory for pictures: Single-trial learning of 2500 visual stimuli. *Psychonomic Science*, 19(2), 73-74.
- Stevens, S.S. (1961). Is there a quantal threshold? In: W.A. Rosenblith (Ed.) *Sensory communication: contributions to the symposium on principles of sensory communication* (p. 806). Cambridge: M.I.T. Press.
- Stevens, S.S. (1975). Psychophysics: introduction to its perceptual, neural, and social prospects. (p. 178). New York: Wiley.
- Stretch, V., & Wixted, J.T. (1998a). Decision rules for recognition memory confidence judgments. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 24(6), 1397-1410.

- Stretch, V., & Wixted, J.T. (1998b). On the difference between strength-based and frequency-based mirror effects in recognition memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *24* (6), 1379-1396.
- Treisman, M., & Williams, T.C. (1984). A theory of criterion setting with an application to sequential dependencies. *Psychological Review*, *91* (1), 68-111.
- Wickens, T.D. (2002). Elementary signal detection theory. (pp. xiii, 262 p.). Oxford ; New York: Oxford University Press.
- Willshaw, D.J., Buneman, O.P., & Longuet-Higgins, H.C. (1969). Non-holographic associative memory. *Nature*, *222* (5197), 960-962.
- Wixted, J.T. (2007). Dual-process theory and signal-detection theory of recognition memory. *Psychological Review*, *114* (1), 152-176.
- Wixted, J.T., & Gaitan, S.C. (2002). Cognitive theories as reinforcement history surrogates: The case of likelihood ratio models of human recognition memory. *Animal Learning & Behavior*, *30* (4), 289-305.
- Wolfe, J.M., Horowitz, T.S., Van Wert, M.J., Kenner, N.M., Place, S.S., & Kibbi, N. (2007). Low target prevalence is a stubborn source of errors in visual search tasks. *Journal of Experimental Psychology: General*, *136* (4), 623-638.
- Yakovlev, V., Amit, D.J., Romani, S., & Hochstein, S. (2008). Universal memory mechanism for familiarity recognition and identification. *Journal of Neuroscience*, *28* (1), 239-248.
- Yonelinas, A.P., & Parks, C.M. (2007). Receiver operating characteristics (ROCs) in recognition memory: a review. *Psychol Bull*, *133* (5), 800-832.