Is working memory inherently more 'precise' than long-term memory? Extremely high fidelity visual long-term memories for frequently encountered objects

Annalise E Miner[1, 2], Mark W. Schurgin[2], Timothy F Brady[2]

[1]Department of Cognitive Science, [2]Department of Psychology
University of California, San Diego, La Jolla, CA, 92093

Please address correspondence to:
Timothy Brady
9500 Gilman Dr. #0109
McGill Hall 5322
La Jolla, CA, 92093
timbrady@ucsd.edu

Word count: 12,123

Abstract

Long-term memory is often considered easily corruptible, imprecise and inaccurate, especially in comparison to working memory. However, most research used to support these findings relies on weak long-term memories: those where people have had only one brief exposure to an item. Here we investigated the fidelity of visual long-term memory in more naturalistic setting, with repeated exposures, and ask how it compares to visual working memory fidelity. Using psychophysical methods designed to precisely measure the fidelity of visual memory, we demonstrate that long-term memory for the color of frequently seen objects is as accurate as working memory for the color of a single item seen 1 second ago. In particular, we show that repetition greatly improves long-term memory, including the ability to discriminate an item from a very similar item ('fidelity'), in both a lab setting (Exps. 1-3) and a naturalistic setting (brand logos, Exp. 4). Overall our results demonstrate the impressive nature of visual long-term memory fidelity, which we find is even higher fidelity than previously indicated in situations involving repetitions. Furthermore, our results suggest that there is no distinction between the fidelity of visual working memory and visual long-term memory, but instead both memory systems are capable of storing similar incredibly high fidelity memories under the right circumstances. Our results also provide further evidence that there is no fundamental distinction between the ‘precision’ of memory and the ‘likelihood of retrieving a memory’, instead suggesting a single continuous measure of memory strength best accounts for working and long-term memory.

Public Significance Statement

Visual working memory appears to be based on persistence of perceptual representations in visual cortex. By contrast, visual long-term memory depends critically on semantically meaningful stimuli and is organized by categories and concepts. Does this mean visual long-term memory is fundamentally incapable of storing as precise perceptual information as visual working memory? In the current work, we show that after being shown multiple repetitions of the same item, visual long-term memory can represent incredibly precise visual details. In fact, after just 8 repetitions, visual long-term memory can be as precise as our very best visual working memories. This provides evidence that there is not a fundamental distinction between the fidelity of visual working memory and visual long-term memory.

Keywords: visual long-term memory; visual working memory; repetition; memory fidelity; memory capacity

Humans have remarkable visual long-term memory abilities, capable of storing thousands of items (Standing, Conezio & Hyber, 1970) with high fidelity (Brady, Konkle, Alvarez & Oliva, 2008). However, while long-term memory can be highly accurate, many researchers have found that it is less accurate than working memory (with claims made in terms of "precision": Biderman et al. 2019; or only "likelihood of retrieval": Brady, Konkle, Gill, Oliva, & Alvarez, 2013) and less robustness to noise (Schurgin & Flombaum, 2018a). For example, Schurgin and Flombaum (2018a) show that adding additional noise to an object image when testing it has almost no effect on working memory but substantially impacts long-term memory, even with identical encoding and test situations. In the current work, we ask (1) whether working memory is truly capable of storing higher fidelity memories than visual long-term memory; (2) whether long-term memories become higher fidelity after repetitions or have an intrinsic limit on the amount of visual detail they can contain; (3) whether multiple distinct processes (e.g., precision errors vs. guesses) are present in working memory and long-term memory, or whether a single process best explains the data; (4) and, ultimately address the question of whether working memory and long-term memory share a representational format or are qualitatively distinct.

Is visual long-term memory less "precise" than working memory?

It may not seem surprising that after a long delay, memory is weaker and correspondingly 'long-term' memories may be less strong and less precise than working memory. This could be true for a variety of reasons: for example, the mere passage of time may particularly impact memory for detail but leave gist unaffected (e.g., Brainerd & Reyna, 2005; Sadeh et al. 2016), or interference may leave categorical knowledge of what we have seen but impair memory for the specific details of individual objects (e.g., Koutstaal & Schacter, 1997; Maxcey & Woodman, 2014). In fact, some researchers argue not only that the two systems tend to differ, but that the visual working memory system is fundamentally different from visual long-term memory system in its ability to represent detailed information (perhaps because working memory necessarily precedes long-term memory, e.g., Biderman et al. 2019).

The idea that working memory is inherently more perceptual than long-term memory is consistent with classic work from the verbal domain showing that working memory interference is based on perceptual similarity but long-term memory interference is based on semantic similarity (Baddeley, 1966). However, in the domain of visual memory, this claim is also partly motivated by the nature of active storage in visual working memory: sensory recruitment models argue that visual working memory arises from persisting perceptual representations in visual cortex (Serences, 2016; Serences, Ester, Vogel, & Awh, 2009; Harrison & Tong, 2009). By their nature, low-level visual representation like this are capable of maintaining significant visual detail. By contrast, long-term memory must necessarily involve consolidated memory representations, likely accessed via medial temporal lobe retrieval structures and so inherently less 'visual' than the case of visual working memory. In fact, many models of the hippocampus and other medial temporal lobe structures argue that a central design feature of this memory system is 'pattern separation' and 'pattern completion' -- designed to group all approximately similar items together into a unified memory representation, and maximize the distinctiveness of this memory from other, similar objects (Yassa & Stark, 2011).

In the case of visual long-term memory, the semantic nature of memory is well known, and broadly consistent with the idea that visual long-term memory may be less perceptual and more semantic than visual working memory. For example, it is known that interference between items in visual long-term memory is based on semantic similarities rather than perceptual overlap (Konkle et al. 2010), and that items interfere with each other within a category-based structure in visual memory (e.g., Maxcey, Glenn, & Stansberry, 2018). Understanding the meaning of a stimulus is also critical to successful encoding into visual long-term memory, as items that are understood are better remembered than identical visual stimuli that are not understood by participants (e.g., Wiseman & Neisser, 1974; Brady, Alvarez & Störmer, 2019). Thus, there are many reasons to suspect that there could be a fundamental difference between working memory and long-term memory in the degree of perceptual detail that can be stored and the tendency to rely on conceptual structure rather than perceptual information.

The role of memory strength in both systems: set size and repetition

However, several important factors are often overlooked when researchers directly compare the precision of representations in these memory systems. One is that working memory is often asked to hold more than just one item in mind simultaneously – for example, to compare two items, we may hold both in mind at once – and because of its limited capacity, this comes with a major cost. In fact, even holding in mind two items rather than one in working memory makes memory for each item far less accurate and precise (e.g., Wilken & Ma, 2004; Zhang & Luck, 2008). Thus, while working memory could be capable of holding more precise estimates of a single item than long-term memory is, with reasonable working memory loads consisting of a few items, it remains possible that working memory represents information with less fidelity than long-term memory.

Another factor that is often overlooked in comparisons between these two systems is that while working memory is necessarily limited to maintaining information that was just present, long-term memory can integrate across many separate episodes. Indeed, in many ways the principle function of long-term memory is to integrate information over time, both to extract categories and other general principles (e.g., Schapiro et al. 2014) as well as to learn about particular objects and how they vary (e.g., Rust & Stocker, 2010). While working memory is designed to work with objects that were just present or that are still present -- and so the object that is the source of the information is straightforward to determine -- long-term memory must connect across large time windows without spatiotemporal cues to what objects are the 'same' as ones that have been previously seen (Schurgin & Flombaum, 2018b). This raises the question of how precise long-term memory can really be: When we have seen a given item many times, is long-term memory at a disadvantage relative to working memory in making detailed discriminations? How accurately can people access existing memory and integrate additional information about an item into these existing long-term memories?

Repetition, spacing and the testing effect

It is well known that long-term memory improves with repetition (e.g., Hintzman, 1976, 2010; Schugin & Flombaum, 2018b), with a large literature demonstrating this for a variety of materials (e.g., pictures: Hintzman & Rogers, 1973; words: Cepeda, Pashler, Vul, Wixted, & Rohrer, 2006), and many influential studies asking about how best to space these repetitions to maximize the improvement in memory (e.g., Cepeda et al. 2006). However, less work has asked about the fidelity of memory (i.e., beyond simply asking whether an item is or is not remembered) and how it is impacted by repetition. Models of memory differ on the extent to which repetition is assumed to independently generate new traces vs. truly integrate new information into higher fidelity memory traces (e.g., Raaijmakers, 2003), and many classic models of memory presume that additional repetition simply increases the probability of retrieval for an item, but does not impact its representational nature (Bower, 1961); for example, arguing new experiences lay down new memory traces rather than integrating with past traces (Hintzman, 1976; Moscovitch et al. 2006; see Kahana, 2020 for a review). Thus, this question is of considerable interest both practically and theoretically.

Work using continuous report measures has provided some mixed evidence on this issue. For example, Sutterer and Awh (2016) asked participants to recreate the color of studied object after a delay using a circular color wheel. For some objects, they gave people retrieval practice. Based on the fit of "mixture models" to their data, which attempt to separate errors into two putatively distinct sources ("the precision of remembered items", and "the proportion of items retrieved successfully"), they argued that retrieval practice seems not to enhance the precision of visual memory (Sutterer and Awh 2016). This is surprising because retrieval practice is among some of the most robust ways to improve memory for items in most situations (e.g., Roediger & Butler, 2011). Thus, this could be taken as evidence that repeated memory traces are not in fact integrated into high fidelity memory traces. However, while Sutterer and Awh (2016) found no effect on "precision", they did find an effect of retrieval practice on the other parameter that the model they fit distinguishes – the proportion of items that were retrieved at all (the opposite of 'guess rate'). Importantly, in several other instances it has been found that with higher power, putative changes in "only" 'proportion of items retrieved, but not item precision, are in fact changes in both (e.g., Zhang & Luck, 2009 vs. Rademaker et al. 2018 in the case of delay). Thus, this work leaves open the possibility that memory fidelity – the accuracy of the memory in terms of the exact color being reproduced -- does in fact improve with repetition and testing practice, not only the ability to access the memory.

Dissociating "precision" from "likelihood of items being retrieved at all"

The majority of existing work asking about the fidelity of visual long-term memory and visual working memory has used methods that attempt to dissociate memory "precision" from the '"likelihood of retrieval" (e.g., in long-term memory: Sutterer & Awh, 2016; Brady et al. 2013; Biderman et al. 2019). However, as noted above, recent work has often empirically found that these two parameters are rarely, if ever, dissociable (that is, higher-power tends to reveal both change, not just one: Rademaker et al. 2018; Biderman et al. 2019). Furthermore, we have recently argued is in fact not even theoretically possible to dissociate 'likelihood of retrieval' from 'precision' in visual memory (Schurgin, Wixted & Brady, 2018). Instead, both of these putatively-distinct parameters seem to tap into single process – characterized by a unitary concept of underlying memory strength, rather than two dissociable psychological constructs (of ‘precision’ and ‘guessing’).

Work claiming a dissociation of these two parameters arose because, when asked to exactly reproduce a color or other aspect of a stimulus from memory, participants often have a substantial number of large errors (a “fat tail” in the error distribution). This is often taken as evidence for a distinct “guessing” or “memory failure” state, the prevalence of which can be estimated via a “mixture model” fit to the data (Zhang & Luck, 2008). However, Schurgin et al. (2018) have recently shown that, counterintuitively, a single process seems to best explain these error distributions –the “fat tail” of errors is just a natural consequence of offering participants many lures that are all maximally dissimilar from a target, not evidence of a distinct “guessing” state (see also Bays, 2014, 2015). The model Schurgin et al. propose is relatively straightforward (see interactive primer at https://bradylab.ucsd.edu/tcc/): if you encode a color (e.g., red), familiarity spreads to other, similar colors (e.g., pink also feels familiar), but not very much to less similar colors (no familiarity spreading to yellow, blue or green). Then, these familiarity signals are corrupted by noise. In this model, memories differ only in their signal-to-noise ratio (d’). The “fat tail” of errors arise because there is almost no spreading of familiarity to any of the colors far from the encoded color (e.g., far from red). Thus, when d’ is low, and thus the noise is high, yellow, blue or green are all equally likely to be the most familiar (due to noise), creating a long, flat tail in the error distribution.

This model (TCC, for Target Confusability Competition) argues that for a given stimulus – for example, a particular color space – there is always a fixed relationship between the so-called “likelihood of retrieval” and “precision” that arise from mixture models, because these both tap the same unitary process, not distinct psychological states. In addition, this model provides a theoretical motivation for believing that repetition of items – which should improve the signal-to-noise ratio (d’) – should not only reduce the likelihood of large errors, but should also improve the “fidelity” of the memory. That is, this model predicts that changing memory strength must, by necessity, not only make people better at easy discriminations (was it red or blue?) but also must improve the fidelity of the memory, improving performance at difficult discriminations (was it light red or dark red?) and continuous report as well.

The current work

Thus, in the current work we sought to address how repetition affects visual long-term memory fidelity. When an item is seen repeatedly, how accurately do people combine the information from each exposure? Does their ability to make subtle perceptual discriminations about the object markedly improve with repetition, or visual long-term memory inherently semantic, and non-visual in nature (e.g., Konkle et al. 2010) in a way that prevents high fidelity visual memories? Does memory fidelity change in the way predicted by single process models of feature memory (e.g., Schurgin et al. 2018) or are there strong dissociations between "precision" and "likelihood of retrieval" (Biderman et al. 2019)?

We addressed these questions using psychophysical methods (in particular, continuous color report), for both newly learned objects with repeated exposure (Experiment 1-3) and existing memory for frequently seen objects in everyday life (brand logos, Experiment 4). In Experiments 1-3 we directly assessed visual long-term memory fidelity for real-world objects in a laboratory setting common to previous studies (Biderman et al. 2019; Brady et al. 2013) but with repeated exposure. In Experiment 4, we used a novel response method where participants selected the color of a previously seen object, not from a color wheel, but from a 2D slice of color space. This allowed us to assess the fidelity of participants color memory for items that do not fall on a single color wheel (i.e. colors used in brand logos).

In all four experiments, we found evidence that participants ability to make very subtle discriminations about the exact color of an object improved a huge amount with additional exposure. In particular, for objects that had been repeatedly seen, participants could accurately reproduce their color as well as they could reproduce the color of a single object held in working memory for just 1 second. In addition, we found that in the tasks using a color wheel, where working memory and long-term memory could be compared directly and where the single process (e.g., TCC) vs. two states (e.g., mixture models) views could be assessed, the two systems had identical error distributions across a wide range of different memory strength conditions – with no dissociations between the supposedly distinct parameters of "precision" and "likelihood of retrieval". Together, these results show that our visual working memory and long-term memory systems do not intrinsically differ in their fidelity; instead, memory strength changes in both systems affect the tendency to make large errors and the precision of small errors in the same way, as would be expected under a single process model of memory (e.g., Schurgin et al. 2018). Furthermore, our results show that long-term memory can be just as high fidelity as our best working memories after repeated exposure, and even standard long-term memory paradigms produce memories with higher fidelity than set-size six working memory.

Experiment 1: The fidelity of visual long-term memory with repeated study

Method

The design, methods, dependent measures and exclusion criteria for this study were preregistered. See: http://aspredicted.org/blind.php?x=3nq63u for details.

Participants: Thirty students were recruited from the University of California, San Diego's undergraduate subject pool, and received class credit for their participation. All subjects gave informed consent, and the study was approved by the University of California, San Diego Institutional Review Board. The sample size was selected a priori (see preregistration) and was considerably larger than the sample sizes used in past literature on this question (e.g., N=5 through N=24; Brady et al. 2013; Biderman et al. 2018). Our main measure of interest was how long-term memory performance benefitted from repetition, e.g., whether LTM performance was improved with 8 repetitions compared to 1. Given the large number of within-participant trials, we expected a large standardized effect size, in line with the difference between working memory and long-term memory performance in previous work, which was very large (all $d_z > 3$, Biderman et al. 2018, 'guess' parameter). Our sample of 30 participants gave us power to detect an effect 1/3rd this size (d_z=1.0) in comparing 8 vs. 1 repetitions with >99% power at alpha=0.05.

Stimuli: 540 object images were selected from a previously published set of stimuli (Brady et al., 2013). These images were of objects in a single arbitrary color (e.g., each object would be recognizable in any color). When presented, each object was colored randomly by rotating the hue of the object on a color wheel, ranging from 0 to 360 degrees. This allowed us to use continuous color report methods to investigate the effect of repetition. Such methods have previously been used to study both working memory for simple shapes, as well as for working and long-term memory of arbitrarily colored object images (Brady et al., 2013; Brady & Alvarez, 2011; Wilken & Ma, 2004; Zhang & Luck, 2008).

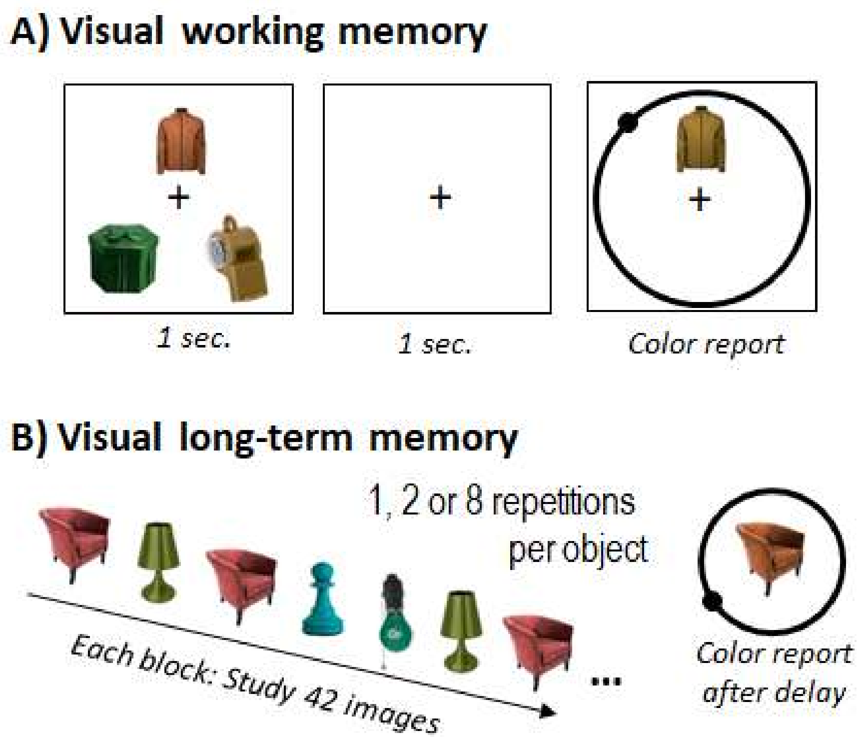

Figure 1. Experiment 1 Methods. *(A) Methods of visual working memory task. Participants saw either 1 or 3 objects for 1 second and had to remember their colors. After a 1 second delay, they used a response wheel to change the color of the object until it matched their memory. (B) Methods of visual long-term memory task. Participants studied 42 images, consisting of some objects seen only once, some repeated twice, and some repeated 8 times. After a delay and a distracting task, participants reported the color of each of these objects using a response wheel.*

Procedure: Participants were asked to remember the precise color of objects and report this color using a color wheel. We compared memory performance in working memory (at set size 1 and 3) and long-term memory (for objects repeated 1, 2 or 8 times). Our primary measure of interest was how memory performance was affected by repetition and to what extent long-term memory performance for well-studied items was comparable to working memory performance.

Overall, participants completed two 1.5-hour experiment sessions with the delay between sessions no more than seven days. In each session, participants completed both a working memory and long-term memory task. These tasks were blocked on each day, with the order counterbalanced across participants and sessions, although the conditions within the working memory task (set size 1, 3) and within the long-term memory task (1 repetition, 2 repetitions, 8 repetitions) were interleaved.

On each trial of the working memory condition, either one or three objects were presented simultaneously for 1 second in a circle around fixation (see Figure 1). Participants were instructed to remember the color of all the presented objects and avoid verbal encoding. After a 1 second delay, participants then reported the color of a randomly chosen object. The probed object appeared in greyscale in the same location it was encoded in, and participants had to match its color to their memory by rotating a response wheel that changed the color of the object. During each session, participants completed 45 trials at each set size, randomly intermixed, for a total of 90 working memory trials. Thus, across both sessions participants completed a total of 180 working memory trials total, 90 trials at each set size.

The long-term memory task was blocked. In each session, participants completed 15 blocks, for a total of 30 blocks across sessions. In each block, participants were shown 42 images, one after another, for 1 second each with a 1 second interstimulus interval. These 42 images were comprised of 6 objects shown only once, 6 objects repeated twice and 3 objects repeated eight times each; images of each object were randomly interleaved in the 42 studied images. Participants were instructed to only remember the color of all the presented objects without using verbal labels, to try to minimize any usage of verbal strategies; importantly previous work by Brady et al. (2013) found nearly no effect of a verbal interference task on this memory task.

A critical aspect of our task was to ensure that participants are not actively storing information in working memory when we are attempting to probe the contents of visual long-term memory. Thus, after the object images were presented, participants completed two trials of a change detection task, to ensure participants weren't actively maintaining colors in visual working memory for the recently encoded items. In the task, adapted from Brady and Tenenbaum (2013), participants were shown a pattern of black and white squares for 750 ms, followed by a 1,000 ms blank period, and then either an identical or changed version of the original display. The test display was shown on screen until participants made a response, indicating if the test was "same" or "different" from the previous display (see Brady & Tenenbaum, 2013, for more information on this task). This filled delay period should disrupt any attempt by participants to actively maintain the colors of studied objects in working memory.

After the change detection task, we assessed long-term memory performance. Participants were asked to report the color of each of the object images that they had previously seen, using the color wheel as in the visual working memory task. An object was cued by being shown in grayscale (at the center of the screen), and then participants had to spin the response wheel to match its color to their memory. In total, during each session participants encoded and were tested on 90 objects that they saw only once; 90 objects they saw repeated twice; and 45 objected repeated eight times. Thus, across both sessions, participants encoded and were tested on 180 objects presented once, 180 presented twice and 90 presented eight times.

Participants were able to complete the experiment at their own pace, without any time constraints or penalties. Participants took on average 1hr 15 minutes for each session. They were instructed to be as accurate as possible.

Data analysis. Participants reported the color of each object image using a color wheel, and therefore the angular difference from the correct answer to the participant's selected answer on the color wheel is our measure of accuracy. On a given trial this error can range from 0° degrees, a perfect memory, to ±180°, a very poor memory. To summarize these errors across trials and estimate overall memory performance, we calculated the deviation of each response in each condition. Then, given a set of responses, we need to compute an overall measure of performance. To do this we relied on the circular standard deviation, which is a descriptive statistic that measures how dispersed participants responses are. This is similar to other descriptive statistics used in the literature (e.g., Bays et al. 2009 report a variant of this; as do van den Berg, Yoo & Ma, 2017 and others, see Ma, 2018).

We use the circular standard deviation in particular as a descriptive statistic of errors because despite being straightforward and non-parametric, it is closely related to model-based measures of performance like *d′* from the single-process Target Confusability Competition model (Schurgin, Wixed & Brady, 2018).

We did not rely primarily on the mixture model technique of Zhang and Luck (2008) because this technique does not appear to isolate different properties of memory (see Introduction and Appendix Section 1); however, our preregistered exclusion criteria did rely upon these mixture model parameters, and we used them for this purpose; they are reported in the Appendix (Section 2).

Our data are well captured by the Target Confusability Competition (TCC) model of Schurgin et al. (2018), providing evidence for this model's generality to long-term memory. These fits are described in the Appendix (Section 1). However, for simplicity – and in line with recommendations for papers not directly about model comparisons (Ma, 2018) -- we report the circular standard deviation as our main measure.

Calculation of chance. If participants had 0 information and simply picked colors at random, their maximum error would be 180 degrees, and their minimum error would be 0 degrees, with a mean of 90 degrees. However, the circular standard deviation of their errors is not the same as their mean error. Thus, to contextualize the circular standard deviations we observe, we calculated chance performance for this metric: To do so, we simply generated 10,000 samples of errors uniformly between -180 and 180, and then calculated the circular standard deviation of this data. This gives us an upper bound on circular standard deviation, indicating what is expected from pure guessing. This is plotted in the Figure 3 as the dashed line.

Exclusion. Following our preregistered exclusion criteria resulted in the exclusion of 6 out of 30 participants. All of these participants were estimated to have a 'guess rate' (Zhang & Luck, 2008) greater than 0.70 in at least one condition. Including these participants did not change the overall pattern of results.

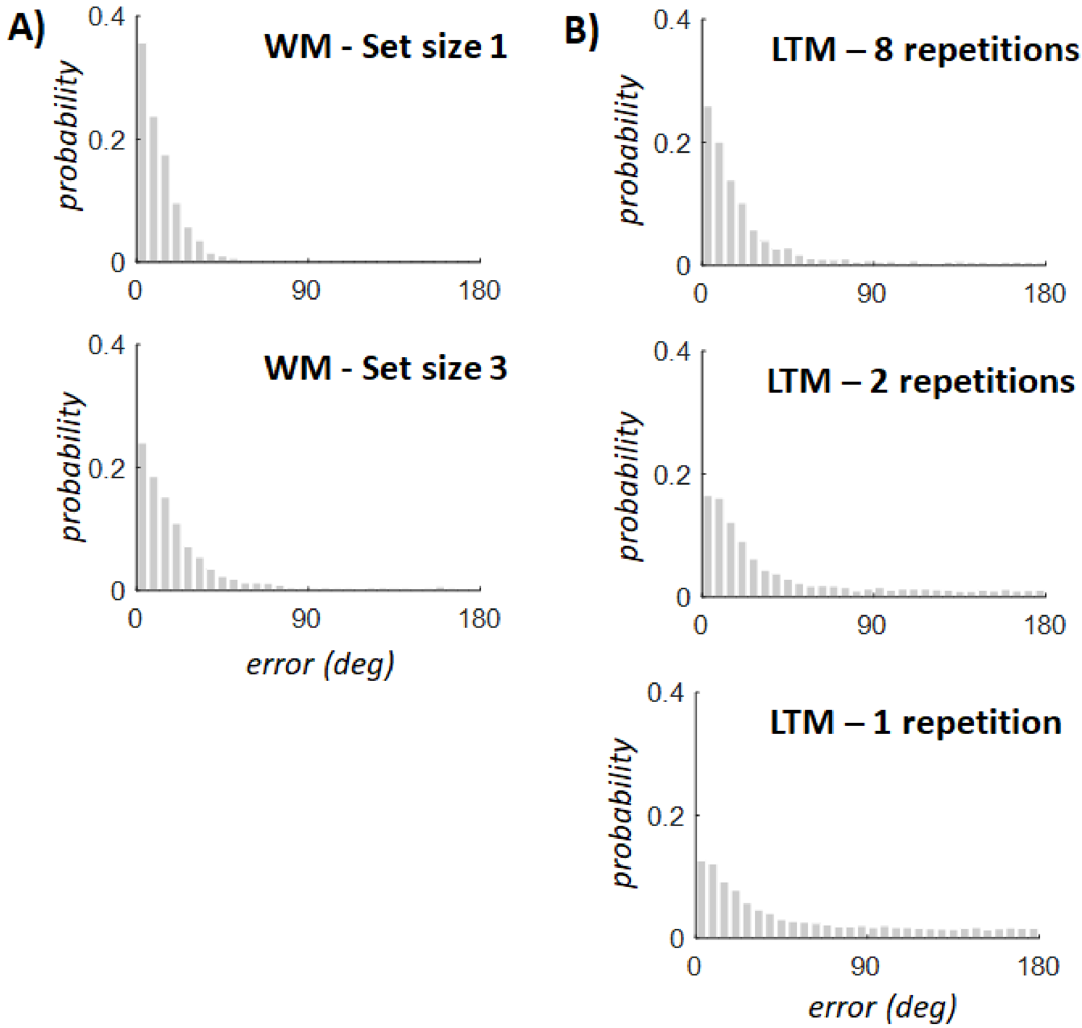

Figure 2. *Error histograms by condition (collapsed across participants for visualization purposes), showing the proportion of each error amount in each condition. 0° error on the reproduction task is perfect memory, and 180° error means participants selected a color on the opposite side of the color wheel. In both working memory and long-term memory, these histograms have the same shape, with many errors near 0 and then a long tail of responses to all the colors that are approximately equally dissimilar to the target. As more items are added in working memory, performance degrades (more large errors); as items repeat more in long-term memory, performance improves (fewer large errors).*

Results

Figure 2 shows error by condition in working and long-term memory, and Figure 3 shows the summary of these errors in terms of circular standard deviation. Overall, we found the expected set size effect in working memory, with performance reliably better for set size 1 than set size 3 ($t(23)=12.1$, $p<0.0001$, $d_z=2.5$). In addition, there was a significant main effect of repetition, with long-term memory performance improving with repetition ($F(2,46)=277.4$, $p<0.0001$, $\eta^2=0.92$).

By 8 repetitions, long-term memory performance was comparable to working memory performance: the circular standard deviation was 28.4 in the 8 repetition case, and 31.2 in the 3 item WM case (difference: -2.8, SEM of difference: 2.1), not reliably different ($t(23)=1.35$, $p=0.19$, $d_z=0.27$). The same results hold when fitting the TCC model to the data (Appendix 1).

Thus, this experiment shows that long-term memory fidelity significantly improves with repetition, even when judged using a psychophysical measurement of exactly what is remembered, and where you must discriminate the remembered item from extremely similar colors. In this situation, long-term memory performance even overlapped with performance in a relatively easy working memory situation: the 8 repetition condition was similar in terms of error to a set size 3 working memory condition.

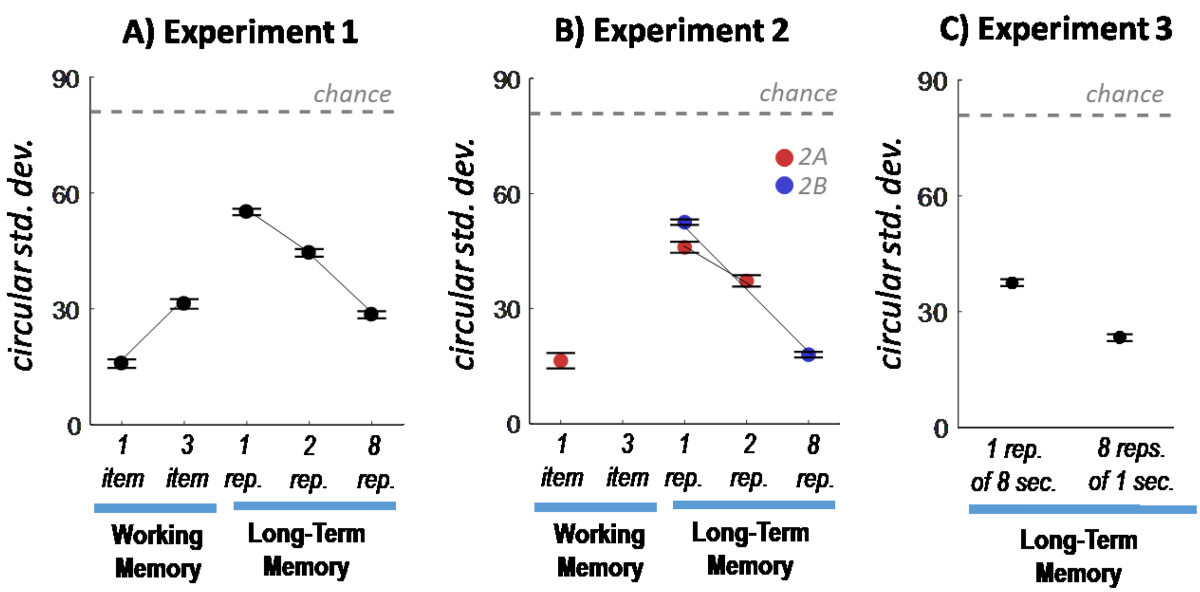

Figure 3. *(A) Results of Experiment 1 in terms of circular standard deviation; each point represents the mean standard deviation across participants, with error bars +/- 1 SEM. As participants were repeatedly exposed to items in long-term memory, memory performance improved. With 8 repetitions in long-term memory, performance was as good as for 3 items that had been seen only 1 second ago (working memory task). (B) Results from the across-subject manipulation in Experiment 2 replicate the within-subject manipulation of Experiment 1. (C) Experiment 3 compared performance for a single 8 sec. exposure in long-term memory to 8 separate 1 sec. exposures, equating total viewing time and asking how memory fidelity is affected. We found participants perform much better with 8 repetitions than a single 8 sec. exposure.*

Experiment 2A and 2B: Across-subject replications of Exp. 1

Experiment 1 provided evidence that participants benefit from repetitions, and showed that under the particular circumstance of our task, 8 repetitions in long-term memory was sufficient to reach the same level of performance as a relatively normal working memory task (with 3 items) – suggesting that the two memory systems are at least partially overlapping in their ability to represent high fidelity color information. In that study, participants by necessity saw the same objects in multiple conditions -- that is, the same object might have appeared in one color in a working memory trial whereas it subsequently appeared in a different color in the long-term memory condition (and was kept constant in the long-term memory condition). Thus, in Experiment 2, we replicated the critical conditions aspects of Experiment 1 in across-subject conditions where objects did not repeat across conditions, to ensure this was not a significant factor.

Method

Experiment 2A: There were N=30 participants (6 excluded per preregistration criterion, final sample: 24). The stimuli, procedure and analysis strategy in Experiment 2A were very similar to those of Experiment 1, but included only a subset of conditions. In particular, in Experiment 2A participants had 3 conditions: (1) perform working memory for 1 item, (2) long-term memory with 1 repetition per item, or (3) long-term memory with 2 repetitions per item. The task was blocked such that participants performed 100 trials of the working memory task either before or after the long-term memory task; and during the long-term memory task, there were 5 blocks, each of 40 images (20 shown once, 10 shown twice in each block).

Experiment 2B: There were N=31 participants (3 excluded per preregistration criterion, final sample: 28). As in Experiment 2A, the stimuli, procedure and analysis strategy in Experiment 2B were very similar to those of Experiment 1, but included only a subset of conditions. In particular, in Experiment 2A participants had only 2 conditions: (1) long-term memory with 1 repetition per item, or (2) long-term memory with 8 repetitions per item. In both conditions, participants saw 24 objects per block. In some blocks participants saw 24 unique objects, whereas in others they saw only 3 objects, each presented 8 times.

Results

Experiment 2A: We found similar results to Experiment 1 in terms of working memory for set size 1 (M: 16.4, SEM: 0.7), long-term memory for unrepeated items (M:46.0, SEM: 3.2) and long-term memory for items repeated twice (M: 37.2, SEM: 3.2). Performance at the long-term memory conditions were significantly worse than the working memory condition (1 repeat: $t(23)=9.26$, $p<0.001$, $d_z=1.89$; 2 repeats: $t(23)=6.42$, $p<0.001$, $d_z=1.31$). The benefit from repetition in long-term memory was also large ($t(23)=4.24$, $p<0.001$, $d_z=0.87$).

Experiment 2B: We found that the circular standard deviation was 52.6 (SEM: 1.8) for items seen once, and 18.0 (SEM: 1.5) for items seen 8 times, a significant difference ($t(27)=-23.96$, $p<0.001$, $d_z=4.53$). In this context, with slightly fewer objects to remember and a blocked design, performance at 8 repetitions was considerably better than in Experiment 1 ($t(50)=4.12$, $p<0.001$, $d=1.15$); in fact, performance was better than the set size 3 working memory task from that experiment (M: 31.2, SEM: 1.6; $t(50)=5.92$, $p<0.001$, $d=1.65$) and numerically not quite as good as set size 1 working memory but comparable statistically (M: 15.8, SEM: 0.6; $t(50)=1.26$, $p=0.21$, $d=0.35$).

Discussion

Experiments 1 and 2 provide strong evidence that long-term memory fidelity significantly improves with repetition. Using a psychophysical measurement of exactly what is remembered, and where you must discriminate the remembered item from extremely similar colors, we found that in the conditions of our task, 8 repetitions of an item brings long-term memory performance to the same level as the best working memory performance (set size 1), with participants able to accurately reproduce the exact color they had seen extremely accurately. That is, 8 repetitions in the long-term memory condition of Experiment 2B allowed people to reproduce the exact color as accurately as they could in the very best working memory conditions of Experiment 1 (one item seen just 1 second ago). This provides evidence that participants do integrate information across repetitions in long-term memory to form higher fidelity memory traces, and provides initial evidence that working memory and long-term memory substantially overlap in the range of fidelity of reproduction that is possible using the two systems, even in situations with nearly maximally strong working memory representations.

Experiment 3: Is repetition better than simply extended encoding time?

Experiments 1 and 2 show that long-term memory is improved dramatically with repetition. Experiment 3 asks whether repetition per se is important, or whether the effect of repetition in those experiments is simply to allow people more total time with each object. Thus, in Experiment 3 we contrast seeing an object and its color 8 times for 1 second each, vs. 1 time for 8 seconds total. If repetition per se has a role in creating higher fidelity memories, than participants should be more accurate in the 8-repetition condition. If total time processing and encoding the objects is most relevant, the two conditions should be identical. And if participants benefit most from a single long exposure, which could potentially allow for deeper processing of the item and its color, then they should be best in the single long exposure condition.

Methods

The design, methods, dependent measures and exclusion criteria for this study were preregistered. See: https://aspredicted.org/blind.php?x=gc8sv2 for details.

Participants: Thirty students were recruited from the University of California, San Diego's undergraduate subject pool, and received class credit for their participation. All subjects gave informed consent, and the study was approved by the University of California, San Diego Institutional Review Board.

Procedure: As in Experiment 1 and 2, participants were asked to remember the precise color of objects and report this color using a color wheel. In this experiment, we probed only long-term memory. We compared memory performance for objects repeated 8 times, shown for 1 sec. each time, to those shown 1 time for 8 seconds.

Each participant completed 24 blocks of study and test. In each block, participants studied 6-48 images, consisting of 6 objects shown either once or 8 times each (48 images). The studied color for each object was randomly chosen by rotating the object in color space, but repeated objects were always shown in the same color each repetition.

After the study period in each block, participants had a filled delay interval designed to disrupt their ability to use visual working memory and ensure we were testing visual long-term memory. In particular, to ensure they could not hold the colors of these images actively in working memory, as in Experiment 1, during the delay participants completed two trials of a change detection task. In the task, adapted from Brady and Tenenbaum (2013), participants were shown a pattern of black and white squares for 750ms, followed by a 1,000ms blank period, and then either an identical or changed version of the original display. The test display was shown on screen until participants made a response, indicating if the test was "same" or "different" from the previous display (see Brady & Tenenbaum, 2013, for more information on this task).

Following this filled delay, they were then probed on the colors of the 6 unique objects using a continuous color wheel, as in Experiments 1 and 2. As in these experiments, we used the circular standard deviation as our main measure of performance.

Exclusion. Data from one participant was lost due to technical error. Following our preregistered exclusion criteria resulted in the exclusion of 0 out of the remaining 29 participants.

Results and Discussion

Although all items were seen for 8 seconds, long-term memory for items repeated 8 times for 1 second each (M: 23.2, SEM: 2.1) was significantly better than long-term memory for unrepeated items shown for 8 seconds (M:37.5, SEM: 2.3; $t(28)=8.02$, $p<0.001$, $d_z=1.49$). This effect was quite large: participants error was nearly halved with 8 separate 1 second exposures compared with a single 8 second exposure. Thus, repetition allows for stronger encoding than does a single presentation of the same amount of exposure.

Thus, repetition is a particularly important tool for forming detailed visual long-term memories. This is consistent with the broadest goal of the visual long-term memory system: integrating information over time, both to extract categories and other general principles (Schapiro et al. 2014) as well as learning about particular objects and how they vary (Rust & Stocker, 2010). While working memory is designed to work with objects that were just present or that are still present, to function effectively, long-term memory must connect across large time windows without spatiotemporal cues to what objects are the 'same' as ones that have been previously seen (Schurgin & Flombaum, 2018b), and repetition and integration across subsequent presentations is a critical aspect of this.

What processes are at work in explaining the repetition benefit? There are several non-mutually exclusive possibilities. One possibility is that re-exposure to an item that has already been seen engages a distinct set of cognitive mechanisms compared to exposure to novel information. For example, people may attempt to recognize it, engaging recognition-specific processes (e.g., Maxcey & Woodman, 2014) including reinstatement of the previous memory trace that allows the new information to be integrated with this previous trace (e.g., Xue, Dong, Chen, Lu, Mumford & Poldrack, 2010). In addition, repeated exposures may cause our memory system to encode slightly differential context each time, leading to more robust memories: classic models of repetition and spacing effects for verbal memory suggest that since memories are inherently contextual, having more varied context of encoding is likely to create more robust memories (e.g., Hintzman, 1974; although see Xue et al. 2010). Finally, it may that repetition allows for stronger encoding simply because people "get more" out of the initial part of any given presentation than the latter part of a presentation (e.g., there is some saturation of how much information is processed as objects remains on the screen; Huebner & Gegenfurtner, 2010). If the majority of the processing of an item happens in the first few hundred milliseconds (e.g., Drugowitsch, Moreno-Bote, Churchland, Shadlen & Pouget, 2012), there will be significant diminishing returns to longer encoding times, but repetition will allow this initial processing to happen repeatedly, resulting in more total information extraction.

Experiment 4: The fidelity of visual long-term memory for brand logos

Can people ever remember items from long-term memory as precisely as they can remember their very best working memories (e.g., which we conceptualize as 1 item seen just a second ago)? Experiment 2 showed one situation where long-term memory for several items – when active maintenance was prevented – was as accurate as participants' very best working memories. However, in that situation the delay time was by necessity short, and the items were very recently encoded and so possibly in a more activated state of long-term memory. Can fully consolidated long-term memories – those most likely to be stored in a non-perceptual format – ever be as accurate as our very best working memories?

To test this, we assessed memory for the color of frequently seen objects - brand logos - as a naturalistic extension of Exp. 1-3. Brand logos are seen in everyday life, and even children show incredibly high recognition rates for logos (Fischer et al., 1991). They are relatively unique in that they are often made up of a single or very few colors, and that there is, at least to a greater extent than most objects, an objective answer to the color they are supposed to be (as opposed to say, the color of an apple or banana – for which there is no truly objective answer). In addition, most logos have been encoded repeatedly over long durations of time (months and years), and, because we do not show the actual color of these logos to participants in the experiment, they thus provide a test of the fidelity of perceptual information in truly long-term, fully consolidated memory.

To test memory for the color of such logos, we collected a set of brand logos that were – based on pilot data – frequently encountered by our participant pool. We then asked participants to both rate their familiarity with these brands and their logos (without seeing them) and then exactly reproduce the color of the logo given only a grayscale version. We then asked participants their confidence in their reproduction.

Because the logos are not all taken from a single circular slice of a color wheel, we cannot directly fit models designed for such data to our data from this experiment (e.g., mixture models; Zhang & Luck, 2008; TCC: Schurgin et al. 2018). However, the insight that there is a single process that explains memory errors even in color wheel data and that simple descriptive statistics of this error therefore do a good job of capturing the relevant factors (i.e., circular standard deviation in Exp. 1-3) means this is not likely to be a significant hurdle to understanding memory in this situation; thus, just as we use the circular standard deviation in Exp. 1-3, we again focus on a simple descriptive statistic of memory error in this Experiment (root mean square error).

Methods

Participants: Thirty students were recruited from the University of California, San Diego's undergraduate subject pool, and received class credit for their participation. All subjects gave informed consent and the study was approved by the University of California San Diego Institutional Review Board. The sample size was selected to match Experiment 1, as similar power is required to again compare the highest familiarity stimuli to the lowest in long-term memory, our main measure of interest. Our post-hoc power in Experiment 1 was even greater than our a priori power calculation took into account, suggesting a similar sample size would again be adequate.

Stimuli: The study consisted of three parts: a working memory color report task, which made use of 140 silhouettes of real-world objects whose color could be completely manipulated (from Sutterer & Awh, 2016; see Figure 2); a long-term memory color report task, using the same stimuli; and a logo color report task. We could not use the object images of Brady et al. (2013) that were used in Experiment 1 because the luminance of these images cannot be manipulated without distorting them, only the hue.

In the brand logo task, participants had to report the exact color of a given brand logo. Thus, to ensure the logos were well known and suitable for our subject population, brand logos were selected via a pilot survey in which UCSD undergraduate participants listed brands for which they could confidently recall a visual memory of the logo. From these responses, we selected brands that were (1) widely reported, and (2) whose most popular logo consisted of largely a single color (excluding black, white and gray). Ultimately, we selected 67 brands, and from their website found their logo and its' dominant color (see stimulus set on OSF: https://doi.org/10.17605/OSF.IO/AQXPN).

Overall Structure of the Experiment: Participants completed three tasks in this experiment (logo memory, long-term memory for newly encoded objects, working memory). Before the first task, participants completed three color perception trials in order to introduce our new color report method. To report colors, in all three tasks, we presented participants with a stimulus on the right side of the screen and a 2D slice of CIELAB space (with fixed L) on the left side of the screen. As participants moved their mouse around the slice of CIELAB space, the color of the relevant part of stimulus on the right side of the screen changed (for silhouettes, this was the entire silhouette; for brand logos, it was only the relevant colored part of the logo). This method allowed participants to report colors not just from a wheel but from an entire slice of color space. The luminance of this slice was always chosen to match the luminance of the correct color; that is, if the correct color was dark, this was a low luminance slice; if the correct color was bright, it was a high luminance slice.

Throughout the experiment, all colors were drawn from the set of colors of the logos. That is, if one of the logos was a particular green, this was the correct answer for that logo in the logo condition; the correct answer for one item in the long-term memory condition; and the correct color for one item in the working memory condition. This ensured that all conditions were comparable, as ultimately the exact same colors were the correct answers and the exact same slices of color space were offered as options in all cases. This is important because, for example, a color that happened to be in the "corner" of the CIE LAB slice will likely elicit a different error distribution than one that happened to be in the middle of the slice. Because of this method of stimulus control, we had participants first complete the brand logo condition (so they would not be pre-exposed to the colors of the logos), then the long-term memory condition, and finally the working memory condition. This was to minimize any potential learning effects of the specific color spaces being used, and, since we were interested in how logo memory compared to working memory, this was the most conservative order (e.g., if any condition would benefit, it would not be logos, but working memory).

Task Procedures: In the first experimental task, the brand logo task, there were 67 trials, one per logo. On each trial, participants were asked to rate their experience with a specific brand logo. Specifically, there were shown the name of the brand (in text; with no logo present), and asked how often they see that logo using a 1-6 scale (1 = Never, 6 = Everyday). After rating their experience with the brand, they then reported the color of the logo. Specifically, the logo appeared in grayscale on the right side of the screen and a fixed L slice of CIELAB color space appeared on the left side of the screen. Using their mouse to hover over the CIELAB color space changed the color of the relevant pixels of the logo on the screen (only those pixels in the color-to-be-reported changed as the mouse moved). Once they had selected a color, participants were asked to rate how confident they were in their choice on a second 6-point scale (1 = Unsure, 6 = Sure).

In the second part of the experiment, participants completed a long-term memory task for newly encoded. Participants were shown 66 object images (taken from Sutterer & Awh, 2016), one after the other for 1 second each with a 1 second interstimulus interval. They were instructed to remember the color of each object image to the best of their ability, without using verbal labels. After the object images were presented, participants completed two trials of a change detection task to ensure participants weren't actively maintaining colors in visual working memory (see Experiment 1 for details). After the change detection task, participants were asked to report the color of each object image that they had seen during the study phase. The color of the object images in the long-term memory task were randomly matched to colors previously used in the brand logo task, such that the exact same colors and exact same slices of CIELAB color space were shown in both tasks.

In the third experimental condition, participants completed a visual working memory task. Participants were shown 1 colored object for one second, and after a one second delay were asked to report the color of the image on the slice of CIELAB color space. They completed 67 working memory trials. Once again, the colors of the object images in this task were randomly matched to specific colors used in both the logo and long-term memory trials of this experiment.

Data Analysis. In this task, the error on each trial is quantifiable as the 2D distance between the correct location on the slice of CIELAB space and the clicked location. As in Experiments 1-3, we use a descriptive statistic of all errors to capture how accurate participants memory is; in particular, we used the root mean squared error (RMSE). The error distribution is significantly skewed, and so to summarize this error for a given participant and condition we use the median RMSE (e.g., the median across all trials of a given condition for a given participant). We then use the mean and standard error of the mean across participants' median's to show the population distribution of medians, since the population distribution of medians is expected to be normally distributed, being itself an aggregate measure.

As noted in the introduction to this Experiment, previous work has largely relied upon circular color report, in which the angular difference between the correct answer and reported answer is taken as the measure of error. This reliance on circular report spaces arose because some models (in particular, mixture models; Zhang & Luck, 2008) claimed to be able to differentiate between different properties of memory using such reports (e.g., precision and likelihood of retrieval). However, as noted, it is now clear that even in circular report spaces, there is really only a single process and thus single parameter being measured (overall memory strength; see Schurgin, Wixted & Brady, 2018). Thus, we believe non-parametric memory error is sufficient to characterize memory both in circular space (Exp. 1-3) and in non-circular space (Exp. 4). However, one drawback of the non-circular space in the current experiment is that chance performance is difficult to characterize. That is, it is unlikely that if people know nothing, they would choose completely at random from the slice (they might avoid corners, for example); and we cannot shuffle responses across trials, since different trials showed different slices of color space. However, since many participants report "Never" having experience with some brand logos, these "1" out of 6 responses on the frequency of experience measure do provide some measure that approximates what chance performance would look like. In addition, a benefit of the 2D approach in the current experiment allows for much more variety in the set of colors shown and tested, allowing us to examine memory for logos and memory for a more realistic range of colors in the working memory and long-term memory conditions.

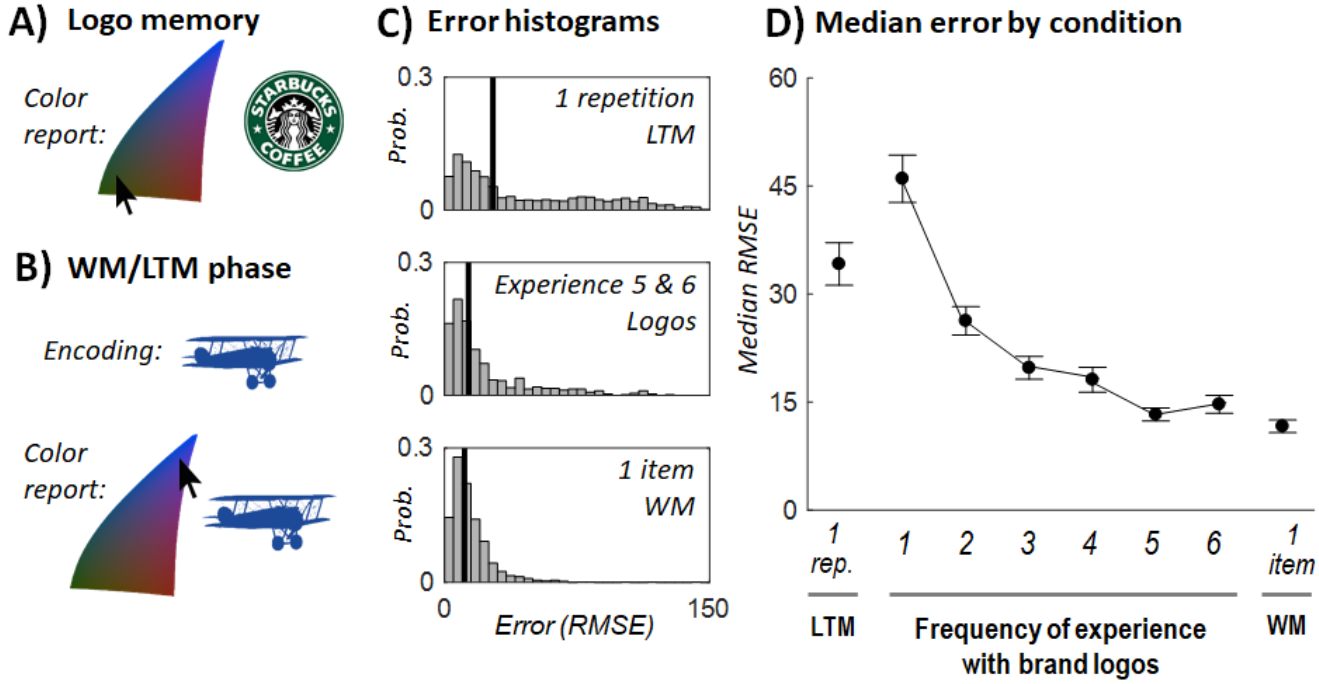

Figure 4. Experiment 3. *(**A**) Methods for the logo memory condition. Participants were given a 2D slice of CIELAB color space, that was matched for the luminance level of the brand logo color. (**B**) Methods for the VWM and LTM conditions. Participants encoded an object silhouette randomly embedded in the same color as one of the logos, and then reported the color of the object at test by clicking the exact color in a 2D slice of CIELAB color space. (**C**) Error histograms across all three conditions, with the median error indicated by a solid black line (**D**) Error by condition. For brand logos, error was calculated as a function of participant's self-reported experience report with that logo (reported before their color memory was tested).*

Results

Figure 4C and 4D show the results across the logo, working memory, and long-term memory conditions. Figure 4C shows the errors collapsed across all participants, showing the full distribution of errors in each condition. This distribution is skewed, with many responses near 0 error and then a fat tail, as is the case in circular color report spaces (Zhang & Luck, 2008; Schurgin, Wixted, & Brady, 2018).

Looking at performance across conditions for each participants (Figure 4D), we find, as expected, that working memory for one item was much more accurate than long-term memory for items that were seen only once ($t(29)=7.3$, $p<0.0001$, $d_z=1.3$). We were primarily interested in how experience with the logos -- as a proxy for stimulus repetition -- affected color memory. Thus, we analyzed the logo data as a function of self-reported experience with the brands. We found that as a participant's self-reported experience with a brand and logo increased, errors in color estimation dramatically decreased, until it was similar to their error for one item seen one second ago (Fig. 4C; $F(5,145)=38.4$ $p<0.0001$, $\eta^2=0.55$).

The logos that participants were least experienced with (1/6; 16.2% of trials) – those they said they'd never seen; effectively a measure of chance performance – were, as expected, reported less accurately than the single-repetition long-term memory items ($t(29)=2.2$, $p=0.04$). The ones they were most experienced with (6/6; 17.1% of trials) were still, on average, quite close to working memory performance even for 1 item, although they were statistically reliably different than the 1 item working memory condition ($t(29)=2.8$, $p=0.01$).

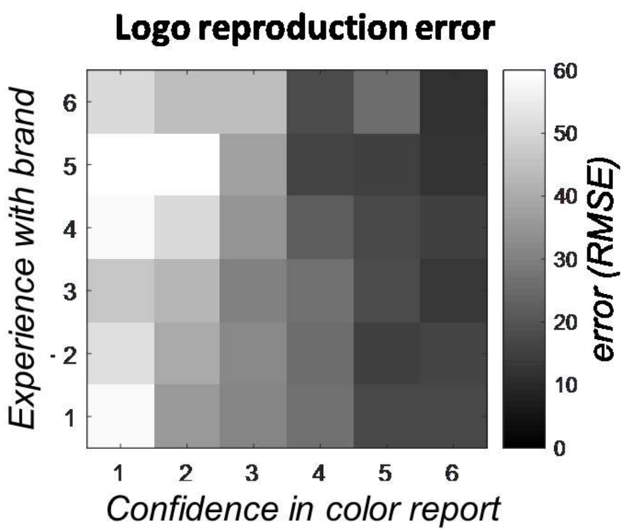

Figure 5. This matrix plots error (RMSE) as a function of reported experience with the brand before being shown any stimulus, and confidence in their color report after. Confidence in color report is strongly related to error (more confident responses had less error), but after taking into account confidence by plotting it separately, there much less of a relationship between reported experience and error. This demonstrates participants had an excellent sense of their own accuracy.

We can also examine memory as a function of self-reported confidence in addition to experience. While these two factors were correlated – people tended to have higher confidence in the color of logos they'd said they had more experience with – they were also somewhat dissociable, with a correlation of r=0.64 (SEM: 0.019) across subjects, corresponding to an $R^2=0.41$. Figure 5 plots error as a function of both variables.

The contrast between Figure 4 – which shows that participants were overall much better when they had more experience with the logo – and Figure 5, where experience seems to play little role in error -- shows this people had an excellent sense of their own accuracy. That is, while people are more likely to report higher confidence when they have more experience, they are approximately equally accurate at a given confidence level regardless of their experience (the dominant structure of Figure 5 is vertical columns, not horizontal stripes). This accurate awareness of their own memory strength means that the major determinant of error in Fig. 5 is confidence, rather than experience. This is consistent with a significant amount of work on "estimator variables" in eyewitness memory (e.g., Semmler, Dunn, Mickes, Wixted, 2018). For example, cross-race identifications tend to be less accurate than same-race identifications. However, the confidence-accuracy relationship is the same for both cross-race and same-race identifications: not only are participants less accurate, but they are also (appropriately) less confident in such identifications. Thus, high confidence reports tend to be equally accurate regardless of estimator variables. Our data support this same conclusion in the case of brand logos.

How did confidence impact memory performance? For logos where people not only reported being extremely experienced with the brand (6/6) but also confident in the color of the logo (6/6), performance (12.3% of trials; median error=11.9) was as good as working memory for an item they had seen 1 second ago (median error=11.6), $t(26)=0.31, p=0.76$; with a Bayes Factor giving 4.7 to 1 evidence in favor of the null hypothesis that these two were equivalent (default JZS Bayes Factor; Rouder, Speckman, Sun, Morey & Iverson, 2009).

Overall, this demonstrates that increased repetition of a brand logo in a naturalistic setting leads to more accurate representations of that logo's color in a participants long-term memory, with the logos people have the most experience with and the most confident memory for being indistinguishable from their memory for an item seen only 1 second before. This is true even given the possible sources of noise in our logo color report task: for example, some brands have changed the color of their logo over time, potentially causing confusions for participants (for example, see: https://www.signs.com/branded-in-memory/); others may have slight differences between the logo color on their website and their real-life signs due to color calibration issues. Nevertheless, despite these sources of noise, brand logo colors were remembered with extremely high fidelity.

General Discussion

Across four experiments, we find that despite the fact that long-term memory is easily corrupted (e.g., Loftus & Palmer, 1996), in the best case scenario where memory is strong and uncorrupted by subsequent interference, long-term memory can be incredibly precise -- a memory for something seen minutes, hours or days ago in the context of many other objects can be as precise as a memory for a single item seen 1 second ago, and accurately discriminated even from very similar colors. This provides strong evidence that participants integrate subsequent exposures into high fidelity memory traces.

Memory for brand logos offers further credence to this claim, as items frequently seen in everyday life were remembered as precisely as the best working memories, despite not having been encountered for hours or days. Critically, this finding may have been obscured if no measure of experience had been collected, as precise logo reports were only observed for items participants reported experiencing regularly and for which they expressed high confidence. Along similar lines, in the study of eyewitness memory high confidence judgments have been shown to be incredibly accurate, contrary to claims that eyewitness memory is unreliable (Wixted & Wells, 2017). Thus, these results provide further evidence that memory strength judgments are critical to understanding the contents of memory.

The fidelity of long-term memory

Humans have remarkable visual long-term memory abilities, capable of storing thousands of items (Standing, Conezio & Hyber, 1970), and previous work has shown that people are extremely good at distinguishing even extremely similar items in visual long-term memory (Brady, Konkle, Alvarez & Oliva, 2008; Hollingworth, 2004, 2005). However, previous work on these lines has largely used meaningful distinctions between objects to test memory (e.g., a full vs. empty mug), preventing a quantitative understanding of memory fidelity.

Recent work looking at visual long-term memory fidelity more quantitatively has often shown worse performance than working memory, both in terms of memory strength or likelihood of retrieval (Biderman et al. 2019; Brady, Konkle, Gill, Oliva, & Alvarez, 2013) and in terms of robustness to noise (Schurgin & Flombaum, 2018a). In some cases, this has been taken as evidence that visual long-term memory is intrinsically lower fidelity than visual working memory (e.g., Biderman et al. 2019), consistent with ideas about neural (e.g., Serences, 2016) and cognitive representation differences between the two systems (Baddeley, 1966) which argue that working memory is inherently more perceptual than long-term memory. However, in the current work, we show that with sufficient repetition, visual long-term memory can be incredibly precise -- people can accurately reproduce nearly the exact color of items they have seen multiple times. This provides evidence that visual long-term memory can be incredibly high fidelity. Thus, despite long-term memory being structured by semantic similarity (e.g., Konkle et al. 2010; Collins & Loftus, 1975), and seemingly relying on an inherently less perceptual neural mechanism of storage (e.g., Serences, 2016), we find that visual long-term memory can store as precise a set of visually detailed information as working memory.

The current work converges with a recent paper by Fresa and Rothen (in press) that showed that in a perceptual learning situation, participants can learn to accurately reproduce colors from visual long-term memory with incredibly high fidelity. In fact, Fresa and Rothen (in press) even showed some degree of generalization, where participants who practiced visual long-term memory color reproduction improved not only at memory for the practiced objects but even at memory for new objects that had been seen only once. This suggests that in addition to repetition improved the fidelity of individual memories, there may be larger scale learning that takes place that affects how accurately people can discriminate items from similar items in memory.

What supports this accurate long-term memory performance? Visual working memories seem to be maintained in visual cortex at least to some extent (Serences, 2016), providing a natural basis for their level of perceptual detail. How can long-term memory have equal detail without such a neural basis for storage? Interestingly, while long-term memory is clearly not actively maintained in perceptual regions, studies have shown that long-term memory retrieval is associated with reaction of the same perceptual brain regions that are activated when perceiving the same items (e.g., Wheeler, Petersen & Buckner, 2000; Kahn, Davachi & Wagner, 2004), with such reinstatment proceeding memory retrieval (e.g., Polyn, Natu, Cohen & Norman, 2005; Xue et al. 2010). Thus, it is possible that visual long-term memory may rely on perceptual regions to access perceptual details in a similar manner to visual working memory, even if it is not actively maintained in these regions.

"Precision" as separate from "likelihood of retrieval": The relationship between working memory and long-term memory

While working and long-term memory are often compared to one another, the majority of research investigating their relative fidelity has been limited to encoding items quickly and just once in long-term memory experiments and comparing this to very strong working memories (e.g., Biderman et al. 2019; Brady et al. 2013; Schurgin & Flombaum, 2018a). The current work provides a suggestion that many of the documented differences between these two systems may not be due to a system-level distinction between them, but rather an artifact of comparing strong working memories to comparatively weak long-term memories.

Indeed, this may explain differences in the results obtained by Brady et al. (2013) and Biderman et al. (2019), who used different set sizes of working memory to draw distinctions between working and long-term memory, with one group arguing for high-fidelity long-term memories and one arguing long-term memory is intrinsically lower fidelity than working memory. Our data lend credence to the idea that working memory and long-term memory are fundamentally similar in representational content, with moderately hard working memory tasks (e.g., set size 3-6) resulting in the same distribution of both similar and dissimilar errors as many long-term memory tasks, and long-term memory tasks with many repetitions giving identical error distributions of similar and dissimilar errors to easy working memory tasks (e.g., set size 1-3).

One way to show this more quantitatively is to compare visual working memory performance to visual long-term memory performance by comparing a number of studies that make use of the same continuous report task using a color wheel. To visualize this, Figure 6 compares visual long-term memory performance from the current set of studies and from past studies to previous data on working memory for color, plotting the parameters of a popular mixture model framework across a wide range of conditions in working memory tasks (see Schurgin et al. 2018 for a similar technique in working memory). This mixture modeling framework takes the distance between the target color and response and models these responses using a mixture model, which attempts to separately quantify memory performance in terms of a 'precision' , and a 'likelihood of retrieval' (or its opposite, a "guess rate"). In the present manuscript, we do not quantify performance in these terms, as it has recently been shown that these parameters are not in fact separable (Schurgin, Wixted & Brady, 2018). Nevertheless, such mixture model parameters are widely reported and provide a window into how accurately participants can discriminate items in memory for similar items. Thus, these parameters allow us to easily compare across memory systems for previous data using the continuous color task. They also allow us to directly compare our data to that of Biderman et al. (2019), who claim that working memory is inherently lower fidelity than long-term memory based on the fits of such model.

In Figure 6, shown in gray are working memory data from a paper that examines many aspects of visual working memory (Schurgin et al., 2018), including performance from set sizes 1-8 and various encoding and delay times. Shown in red are the data from the long-term memory color report tasks of both Biderman et al. (2019) and the current manuscript. As can be clearly seen, the two parameters trade-off nearly identically in the two memory systems, with the curves completely overlapping. In fact, the lowest performance – in terms of both "guess rate" and precision (SD) – comes from the working memory conditions (set size 6 and 8), where people are less accurate than in any of the long-term memory conditions tested in the current paper or by Biderman et al. (2019). Thus, contrary to Biderman et al. (2019), we do not observe any evidence in favor of the idea that long-term memory has intrinsically lower fidelity than working memory (e.g., noisier representations, with larger standard deviations). Instead, our data show that if you compare a wide range of standard long-term and working memory tasks, you find identical data distributions and parameters that fit those distributions.

Importantly, our data also reveal that no individual points alone are sufficient to understand the relationship between the fidelity of these two systems, as memory strength can vary greatly in both systems. Biderman et al. (2019) compared long-term memory data only to set size 3 working memory, thus finding a working memory advantage; if they had compared long-term memory instead to set size 6 working memory data, they would have found a long-term memory advantage. Only by plotting a wide range of memory strengths together does it become clear that the two systems lie on the same curve.

Figure 6. *Gray circles indicate data from visual working memory for color across a range of set size (1-8), encoding times and delays from Schurgin, Wixted & Brady (2018). Unfilled circles come from set sizes 1 and 3; filled gray circled come from set size 6 and 8. Red circles are data from the long-term memory conditions of Biderman et al. (2019); red diamonds are data from the current paper. The black line represents the prediction of the Schurgin et al. (2018) TCC model, which argues that both parameters derive from a single process rather than being dissociable psychological components The tight coupling of the two parameters ("guess rate" and "SD") across a wide range of conditions is strongly consistent with the idea that the parameters of the mixture model reflect one process, not two (as separately shown by Schurgin et al. 2018). The red LTM points falling on the same line as gray WM ones provides evidence that this coupling is the same for working memory and long-term memory. Note that the long-term memory conditions in both the current paper and Biderman et al. (2019) are both better – down and to the left – than the set size 6 and 8 working memory conditions (filled gray circle), and several conditions in the current paper are as accurate as even the best working memory conditions observed in Schurgin et al. (2018).*

It is also important to note that the strong relationship observed between the 'guess' and 'precision' parameters in both the working memory and long-term memory data converge with the proposal from Schurgin et al. (2018) that these parameters are not distinct, but tap just a single underlying process. The plot in Figure 6 is a state-trace plot (Dunn & Kalish, 2018), and is completely consistent with a single process model – where "precision" and "likelihood of retrieval" are just two ways of measuring the same underlying variable (memory strength). Furthermore, the black dashed line in Figure 6 is the prediction of the Target Confusability Competition (TCC) model proposed by Schurgin et al. (2018) – this model says that by necessity, when using this color space, the only possible mixture model parameters that can arise are the ones on that line (subject to measurement error). The current long-term memory data are clearly consistent with this prediction. Thus, the current data also provide additional evidence there is effectively only a single parameter of memory difficulty observed in continuous reproduction error histograms.

How should we think about the "precision" of working memory vs. long-term memory in this framework? The TCC model, consistent with the state-trace plot (Fig. 6), suggests that there is no such concept as the 'precision' of a memory system. Instead, there is only a concept of 'memory strength', which combines with a fixed similarity function for a given stimulus space (see https://bradylab.ucsd.edu/tcc/). The way this memory strength manifests in terms of the errors people make, and in terms of their ability to make discriminations between similar vs. dissimilar items, appears to be the same for working memory and long-term memory. However, the stimulus space matters quite a bit: that is, different stimuli spaces (e.g., different color wheels, or different features) have different characteristic similarity functions, and thus different shaped error distributions and different mixture model parameter (Schurgin et al. 2018). Thus, rather than the difficulty of discriminating items from similar ones arising due to differential limits in the 'precision' memory systems (e.g., Biderman et al. 2019), these limits seem to result from differences in the underlying similarity structure of the perceptual dimensions being studied (e.g., the color wheel being used). Overall, then, our data suggest that visual working memory and visual long-term memory largely overlap in their ability to represent high fidelity color information – either in terms of mixture model parameters (Fig. 6) or simple descriptive statistics of error. Thus, difficult long-term memory tasks and difficult working memory both result in the same "standard deviation" and same "guess rate"; easy working memory and easy long-term memory tasks likewise result in identical memory parameters. This suggests that not only can long-term memory hold precise memories but that memory fidelity functions similarly in the two memory systems.

Are working memory and long-term memory the same system?

There is significant evidence for shared principles between working memory and long-term memory, particularly for verbal stimuli (Jonides et al., 2008; McElree, 2006; Nairne, 2002). For example, items putatively held in active storage are not accessed any faster than those held in passive storage (McElree, 2006), and both systems can be integrated in some temporal context views of memory (Brown, Neath & Chater, 2007). Similarly, there appear to be shared principles of access and refreshing between working memory and long-term memory (e.g., Ranganath, Johnson & D'Esposito, 2003), resulting in some claims that there may be no need to posit two distinct memory systems (Ranganath & Blumenfeld, 2005).

The current work is consistent with another important way in which working memory and long-term memory are not distinct: representations in both systems appear to have the same fidelity, and, indeed, asking participants to reproduce colors in both systems not only produces similar distributions, but seemingly identical ones, both in terms of the "heavy tail" and the width of the central part of the distribution (Figure 6). Does this mean working memory and long-term memory are not in any way distinct?

We find the evidence from neuroscience that there are different processes going on when accessing actively maintained information vs. passively stored information compelling. For example, there is clear continued firing in the form of the Contralateral Delay Activity (Vogel & Machizawa, 2004) when participants actively maintain color information in working memory, but this is not present if the information has already been stored in long-term memory (e.g.,Carlisle, Arita, Pardo & Woodman, 2011). fMRI evidence also strongly suggests active storage during the working memory delay for visual stimuli (e.g., Xu & Chun, 2006; Harrison & Tong, 2009). Similarly, hippocampal damage seems to, at least in some instances, selective impair long-term memory access but not working memory access, particularly for small numbers of items (e.g., Jeneson & Squire, 2012). How can these ideas – on one hand, evidence for a unified system, with similar fidelity; and on the other hand, clearly distinct and more active neural substrates for working memory -- be reconciled?

One possibility with significant support in the literature is that working memory and long-term memory are different processes for working with the same underlying memory representations. That is, while the representations are themselves the same, it is possible to keep these representations actively accessible with attention – "working memory" – or to allow them to become passive, and then retrieve them later ("long-term memory"). These different ways of working with memories are importantly distinct, but the memories themselves may not be. This is broadly consistent with the view of working memory as "activated" long-term memory representations (e.g., Cowan, 1999; Lewis-Peacock & Postle, 2008).

Conclusion

We show that repetition, either in the lab or naturalistically, leads to incredibly high fidelity long-term memories, such that items with which we have significant experience can be reproduced in a continuous report task as accurately as if they had just been seen. In particular, with more repetitions, people are able to accurately reproduce a color extremely precisely -- as precisely as an item seen one second ago in visual working memory. Despite the fundamentally different neural substrate of visual working memory, with items stored and maintained in perceptual regions (e.g., Serences, 2016), visual working memory does not seem to have an intrinsic advantage in making fine-grained discriminations compared to visual long-term memory. Instead, memory strength – which varies a large amount in both working memory and long-term memory -- is the main driver of the ability to make fine-grained judgments about the exact perceptual features of previously seen objects, independent of memory system.

Author Contributions

A. Miner and T. Brady developed the study concepts. All authors contributed to the study design. Testing and data collection were performed by A. Miner. A. Miner, M. Schurgin and T. Brady performed the data analysis and interpretation. A. Miner and T. Brady drafted the manuscript, and M. Schurgin provided critical revisions. All authors approved the final version of the manuscript for submission.

Appendix

1. Target Confusability Competition (TCC) model fits to the long-term memory data

Previous research comparing the fidelity of color memory across working and long-term memory, such as Brady et al. (2013) and Biderman (2019), relied on mixture models, which quantify memory performance in terms of two putatively distinct concepts: a ‘precision’ (strength of information in memory), and a ‘guess rate’ (probability an item is in memory) (Zhang & Luck, 2008).

In the present manuscript, we do not quantify performance in these terms, as it has recently been shown that these parameters are not in fact separable (Schurgin, Wixted & Brady, 2018). That is, large errors – which result in a “long tail” often interpreted as evidence of discrete guessing – appear to arise from the same process as do small errors. In light of this finding, separately modeling different aspects of memory is unnecessary -- an item’s memory strength can be quantified in signal detection terms as d′ (Schurgin, Wixted & Brady, 2018) or non-parametrically (e.g., using the circular standard deviation of participants’ errors) -- but in either case, there appears to be no separate process of ‘guessing’ that needs to be accounted for. Thus, to summarize errors across trials and estimate overall memory performance in the present manuscript, we calculated the circular standard deviation of responses by condition. The circular standard deviation (sometimes known as the angular deviation) has been recommended as a measure because despite being straightforward and non-parametric, it is closely related to model-based measures like *d′* (Schurgin, Wixed & Brady, 2018).

However, rather than simply using the circular standard deviation, it is also possible to fit the Target Confusability Competition (TCC) model to the data from the experiments we do with continuous report, to obtain *d’,* a measure of memory strength. Doing so reveals that the model accurately fits both the working memory and long-term memory data, and gives substantially similar conclusions to the circular standard deviation analyses (e.g., Figure A1).

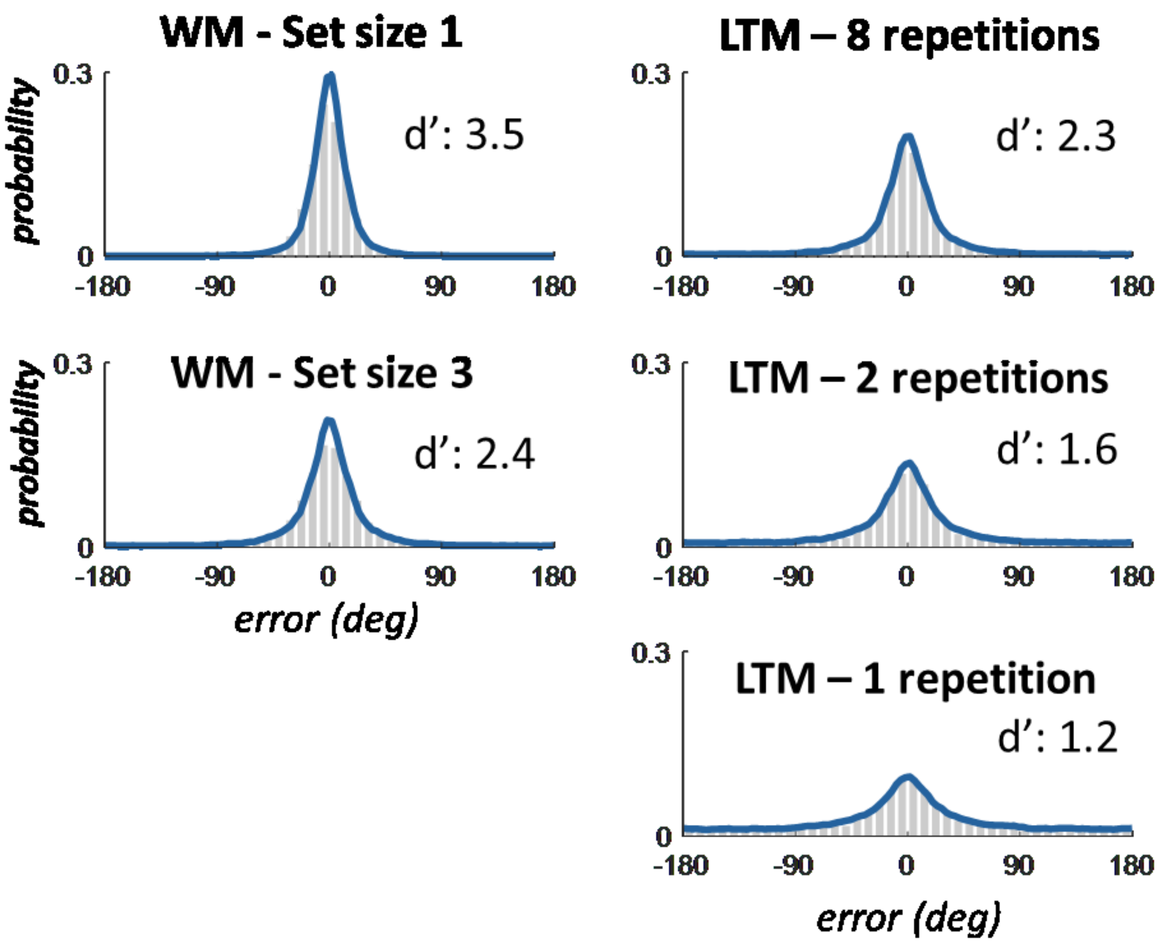

Figure A1. *Fits of TCC to Experiment 1 data. Blue is the fit of the 1-parameter (d') TCC model that assumes a single process generates all errors (e.g., that there is no discrete guess state). Gray is the histogram of participants errors. The d' values are the fit to the data collapsed across all participants; the average and variation in d' across participants for all experiments is reported in the table below. Note that the d' of the fit to the average data is not the same as the average d' of fits to individual subjects.*

The average and SEM of the memory strength (d') values for each condition are reported below.

Experiment 1

	WM - set size 1	*WM - set size 3*	*LTM - 1 repeat*	*LTM - 2 repeats*	*LTM - 8 repeats*
d'	3.73 (0.09)	2.54 (0.10)	1.42 (0.11)	1.94 (0.12)	2.81 (0.15)

Experiment 2A

	WM - set size 1	*LTM - 1 repeat*	*LTM - 2 repeats*
d'	3.70 (0.12)	2.53 (0.15)	3.90 (0.08)

Experiment 2B

	LTM - 1 repeat	*LTM - 8 repeats*
d'	1.52 (0.09)	3.73 (0.14)

Experiment 3

	LTM - 1 repeat of 8 seconds	*LTM - 8 repeats of 1 seconds*
d'	2.29 (0.13)	3.30 (0.17)

2. Mixture model fits for Experiments 1, 2 and 3

Although we no longer have reason to believe that previously reported evidence supports the mixture model's distinction between two separate aspects of memory (number of items; precision of those items), our pre-registered analysis plan suggested the use of not only the non-parametric angular deviation but also mixture model parameter estimates. Thus, we report the mixture model parameters here for Experiment 1-3. Note that they are consistent with the claim we make using non-parametric methods: both in terms of guess rate and standard deviation, repetition improved long-term memory, and 8 repetitions improves performance to approximately the level of set size 3 working memory. Data is formatted as mean (SEM).

Experiment 1

	WM - set size 1	*WM - set size 3*	*LTM - 1 repeat*	*LTM - 2 repeats*	*LTM - 8 repeats*
guess	0.01 (0.003)	0.11 (0.02)	0.44 (0.03)	0.29 (0.03)	0.10 (0.02)
SD	15.3 (0.4)	20.6 (0.9)	23.8 (1.4)	21.4 (1.1)	18.3 (0.97)

Experiment 2A

	WM - set size 1	*LTM - 1 repeat*	*LTM - 2 repeats*
guess	0.01 (0.003)	0.29 (0.04)	0.21 (0.05)
SD	15.2 (0.7)	24.9 (1.6)	19.9 (1.4)

Experiment 2B

	LTM - 1 repeat	*LTM - 8 repeats*
guess	0.39 (0.03)	0.04 (0.009)
SD	23.8 (2.0)	13.9 (0.7)

Experiment 3

	LTM - 1 repeat of 8 seconds	*LTM - 8 repeats of 1 seconds*
guess	0.20 (0.03)	0.07 (0.02)
SD	20.6 (1.1)	17.0 (1.1)

They also show, as previously reported by Schurgin et al (2018) and visualized in Figure 6, a strong relationship between SD and guess estimates, consistent with the idea that they vary along a single dimension and in fact reflect the outcome of only a single process (see General Discussion).

3. Replication of Experiment 2

In another experiment (Experiment S1), participants performed only the long-term memory task, for items repeated either once or 8 times. N=33 participants (7 excluded per preregistration criterion, final sample: 26) saw 24 objects per block. The task was blocked such that in some blocks participants saw 24 unique objects, whereas in others they saw only 3 objects, each presented 8 times. Immediately following the last object (e.g., with no change detection task), we found that angular deviation was 51.8 (SEM: 2.2) for items seen once, and 16.6 (SEM: 1.1) for items seen 8 times. This is consistent with the results of Experiment 2B.

References

Baddeley, A. D. (1966). Short-term memory for word sequences as a function of acoustic, semantic and formal similarity. *Quarterly journal of experimental psychology*, 18(4), 362-365.

Bays, P. M., Catalao, R. F., & Husain, M. (2009). The precision of visual working memory is set by allocation of a shared resource. *Journal of vision*, 9(10), 7-7.

Bays, P. M. (2014). Noise in neural populations accounts for errors in working memory. *Journal of Neuroscience*, 34(10), 3632-3645.

Bays, P. M. (2015). Spikes not slots: noise in neural populations limits working memory. *Trends in cognitive sciences*, *19*(8), 431-438.

Biderman, N., Luria, R., Teodorescu, A. R., Hajaj, R., & Goshen-Gottstein, Y. (2019). Working Memory Has Better Fidelity Than Long-Term Memory: The Fidelity Constraint Is Not a General Property of Memory After All. *Psychological Science*, 30(2), 223-237.

Bower, G. H. (1961). Application of a model to paired-associate learning. *Psychometrika*, *26*(3), 255-280.

Brady, T. F., & Alvarez, G. A. (2011). Hierarchical encoding in visual working memory: Ensemble statistics bias memory for individual items. *Psychological science*, 22(3), 384-392.

Brady, T. F. and Alvarez, G.A. (2015). Contextual effects in visual working memory reveal hierarchically structured memory representations. *Journal of Vision*, 15(15):6.

Brady, T. F., Alvarez, G., and Störmer, V. (2019). The role of meaning in visual memory: Face-selective brain activity predicts memory for ambiguous face stimuli. *Journal of Neuroscience*, 39 (6) 1100-1108.

Brady, T. F., & Tenenbaum, J. B. (2013). A probabilistic model of visual working memory: Incorporating higher order regularities into working memory capacity estimates. *Psychological Review*, *120*(1), 85.

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

Brady, T. F., Konkle, T.F., Gill, J., Oliva, A. and Alvarez, G.A. (2013). Visual long-term memory has the same limit on fidelity as visual working memory. *Psychological Science*, 24(6), 981-990.

Brainerd, C. J., & Reyna, V. F. (2005). *The science of false memory* (Vol. 38). Oxford University Press.

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

Carlisle, N. B., Arita, J. T., Pardo, D., & Woodman, G. F. (2011). Attentional templates in visual working memory. Journal of Neuroscience, 31(25), 9315-9322.

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.

Collins, A. M., & Loftus, E. F. (1975). A spreading-activation theory of semantic processing. *Psychological review*, 82(6), 407.

Cowan, N. (1999) An embedded-processes model of working memory. In *Models of Working Memory* (Miyake, A. and Shah, P., eds), pp. 62–101, Cambridge University Press

Drugowitsch, J., Moreno-Bote, R., Churchland, A. K., Shadlen, M. N., & Pouget, A. (2012). The cost of accumulating evidence in perceptual decision making. *Journal of Neuroscience*, *32*(11), 3612-3628.

Dunn, J. C., & Kalish, M. L. (2018). *State-trace analysis.* Springer.

Fischer, P. M., Schwartz, M. P., Richards, J. W., Goldstein, A. O., & Rojas, T. H. (1991). Brand logo recognition by children aged 3 to 6 years: Mickey Mouse and Old Joe the Camel. *Jama*, *266*(22), 3145-3148.

Fresa, R.O., & Rothen, N. (in press). Training Enhances Fidelity of Color Representations in Visual Long-Term Memory. *Journal of Cognitive Enhancement.*

Harrison, S. A., & Tong, F. (2009). Decoding reveals the contents of visual working memory in early visual areas. *Nature*, 458(7238), 632.

Hintzman, D. L. (1974). Theoretical implications of the spacing effect.

Hintzman, D. L., & Rogers, M. K. (1973). Spacing effects in picture memory. *Memory & Cognition*, *1*(4), 430-434.

Hintzman, D. L. (1976). Repetition and Memory. In *Psychology of learning and motivation* (Vol. 10, pp. 47-91). Academic Press.

Hintzman, D.L. (2010) How does repetition affect memory? Evidence from judgments of recency. *Memory & Cognition*, 38, 102.

Hollingworth, A. (2004). Constructing visual representations of natural scenes: the roles of short- and long-term visual memory. *Journal of Experimental Psychology: Human Perception and Performance*, 30(3), 519.

Hollingworth, A. (2005). The relationship between online visual representation of a scene and long-term scene memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 31(3), 396.

Huebner, G. M., & Gegenfurtner, K. R. (2010). Effects of viewing time, fixations, and viewing strategies on visual memory for briefly presented natural objects. *The Quarterly Journal of Experimental Psychology*, *63*(7), 1398-1413.

Jeneson, A., & Squire, L. R. (2012). Working memory, long-term memory, and medial temporal lobe function. Learning & memory, 19(1), 15-25.

Jonides, J., Lewis, R. L., Nee, D. E., Lustig, C. A., Berman, M. G., & Moore, K. S. (2008). The mind and brain of short-term memory. *Annu. Rev. Psychol.*, 59, 193-224.

Kahana, M. J. (2020). Computational Models of Memory Search. Annual Review of Psychology, 71, 107-138.

Kahn, I., Davachi, L., & Wagner, A. D. (2004). Functional-neuroanatomic correlates of recollection: implications for models of recognition memory. Journal of Neuroscience, 24(17), 4172-4180.

Konkle, T., Brady, T. F., Alvarez, G. A., & Oliva, A. (2010). Conceptual distinctiveness supports detailed visual long-term memory for real-world objects. Journal of Experimental Psychology: General, 139(3), 558.

Koutstaal, W., & Schacter, D. L. (1997). Gist-based false recognition of pictures in older and younger adults. *Journal of memory and language*, *37*(4), 555-583.

Lewis-Peacock, J. A., & Postle, B. R. (2008). Temporary activation of long-term memory supports working memory. *Journal of Neuroscience*, *28*(35), 8765-8771.

Loftus, E. F., & Palmer, J. C. (1996). Eyewitness testimony. In *Introducing psychological research* (pp. 305-309). Palgrave, London.

Ma, W. J. (2018). Problematic usage of the Zhang and Luck mixture model. *Biorxiv*, 268961.

Maxcey, A. M., & Woodman, G. F. (2014). Forgetting induced by recognition of visual images. *Visual cognition*, *22*(6), 789-808.

Maxcey, A. M., Glenn, H., & Stansberry, E. (2018). Recognition-induced forgetting does not occur for temporally grouped objects unless they are semantically related. *Psychonomic Bulletin & Review*, 25(3), 1087-1103.

McElree, B. (2006). Accessing recent events. *Psychology of learning and motivation*, 46, 155-200.

Moscovitch M, Nadel L, Winocur G, Gilboa A, Rosenbaum RS. 2006. The cognitive neuroscience of remote episodic, semantic and spatial memory. Curr. Opin. Neurobiol. 16:179–90

Nairne, J. S. (2002). Remembering over the short-term: The case against the standard model. *Annual review of psychology*, 53(1), 53-81.

Polyn, S. M., Natu, V. S., Cohen, J. D., & Norman, K. A. (2005). Category-specific cortical activity precedes retrieval during memory search. *Science*, *310*(5756), 1963-1966.

Raaijmakers, J. G. (2003). Spacing and repetition effects in human memory: Application of the SAM model. *Cognitive Science*, *27*(3), 431-452.

Rademaker, R. L., Park, Y. E., Sack, A. T., & Tong, F. (2018). Evidence of gradual loss of precision for simple features and complex objects in visual working memory. *Journal of Experimental Psychology: Human Perception and Performance*, *44*(6), 925.

Ranganath, C., & Blumenfeld, R. S. (2005). Doubts about double dissociations between short- and long-term memory. *Trends in cognitive sciences*, 9(8), 374-380.

Ranganath, C., Johnson, M. K., & D'Esposito, M. (2003). Prefrontal activity associated with working memory and episodic long-term memory. *Neuropsychologia*, 41(3), 378-389.

Roediger, H. L., & Butler, A. C. (2011). The critical role of retrieval practice in long-term retention. *Trends in cognitive sciences*, 15(1), 20-27.

Rouder, J. N., Speckman, P. L., Sun, D., Morey, R. D., & Iverson, G. (2009). Bayesian t tests for accepting and rejecting the null hypothesis. *Psychonomic Bulletin & Review*, *16*(2), 225-237.

Rust, N. C., & Stocker, A. A. (2010). Ambiguity and invariance: two fundamental challenges for visual processing. *Current opinion in Neurobiology*, 20(3), 382-388.

Sadeh, T., Ozubko, J. D., Winocur, G., & Moscovitch, M. (2016). Forgetting patterns differentiate between two forms of memory representation. *Psychological science*, *27*(6), 810-820.

Schapiro, A.C., Gregory, E., Landau, B., McCloskey, M., Turk-Browne, N.B. (2014). The necessity of the medial temporal lobe for statistical learning. *Journal of Cognitive Neuroscience,* 26:8, 1736-1747.

Schurgin, M. W., & Flombaum, J. I. (2018). Visual Working Memory is More Tolerant Than Visual Long-Term Memory. *Journal of Experimental Psychology: Human Perception and Performance, 44*(8), 1216-1227.

Schurgin, M. W., & Flombaum, J. I. (2018). Properties of Visual Episodic Memory Following Repeated Encounters with Objects. *Learning & Memory, 25*(7), 309-316.

Schurgin, M. W., Wixted, J. T., & Brady, T. F. (2018). Psychophysical Scaling Reveals a Unified Theory of Visual Memory Strength. *BioRxiv*, 325472.

Semmler, C., Dunn, J., Mickes, L., & Wixted, J. T. (2018). The role of estimator variables in eyewitness identification. *Journal of Experimental Psychology: Applied*, 24(3), 400.

Serences, J. T. (2016). Neural mechanisms of information storage in visual short-term memory. *Vision research*, 128, 53-67.

Serences, J. T., Ester, E. F., Vogel, E. K., & Awh, E. (2009). Stimulus-specific delay activity in human primary visual cortex. *Psychological science,* 20(2), 207-214.

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.

Sutterer, D. W., & Awh, E. (2016). Retrieval Practice Enhances the Accessibility but not the Quality of Memory. *Psychonomic Bulletin & Review*, *23*(3), 831–841.

Van den Berg, R., Yoo, A. H., & Ma, W. J. (2017). Fechner's law in metacognition: A quantitative model of visual working memory confidence. *Psychological review*, 124(2), 197.

Vogel, E. K., & Machizawa, M. G. (2004). Neural activity predicts individual differences in visual working memory capacity. *Nature*, *428*(6984), 748.

Wheeler, M. E., Petersen, S. E., & Buckner, R. L. (2000). Memory's echo: vivid remembering reactivates sensory-specific cortex. *Proceedings of the National Academy of Sciences*, 97(20), 11125-11129.

Wilken, P., & Ma, W. J. (2004). A detection theory account of change detection. *Journal of Vision*, *4*(12), 11-11.

Wiseman, S., & Neisser, U. (1974). Perceptual organization as a determinant of visual recognition memory. *The American Journal of Psychology*, 87(4), 675–681.

Wixted, J. T. & Wells, G. L. (2017). The Relationship between Eyewitness Confidence and Identification Accuracy: A New Synthesis. *Psychological Science in the Public Interest, 18*, 10-65.

Xu, Y., & Chun, M. M. (2006). Dissociable neural mechanisms supporting visual short-term memory for objects. *Nature*, 440(7080), 91-95.

Xue, G., Dong, Q., Chen, C., Lu, Z., Mumford, J. A., & Poldrack, R. A. (2010). Greater neural pattern similarity across repetitions is associated with better memory. *Science*, *330*(6000), 97-101.

Yassa, M. A., & Stark, C. E. (2011). Pattern separation in the hippocampus. *Trends in Neurosciences*, *34*(10), 515-525.

Zhang, W., & Luck, S. J. (2008). Discrete fixed-resolution representations in visual working memory. *Nature*, *452*, 233–235.

Zhang, W., & Luck, S. J. (2009). Sudden death and gradual decay in visual working memory. *Psychological science*, *20*(4), 423-428.