Visual long-term memory has the same limit on fidelity as visual working memory

Timothy F. Brady\textsuperscript{1}, Talia Konkle\textsuperscript{1}, Jonathan Gill\textsuperscript{2}, Aude Oliva\textsuperscript{2,3}, & George A. Alvarez\textsuperscript{1}

\textsuperscript{1}Department of Psychology
Harvard University

\textsuperscript{2}Department of Brain & Cognitive Sciences
\textsuperscript{3}Computer Science and Artificial Intelligence Laboratory
Massachusetts Institute of Technology

Please address correspondence to:
Timothy Brady
Department of Psychology
Harvard University
33 Kirkland St.
Cambridge, MA 02138
Email: tbrady@wjh.harvard.edu

Running Head: Fidelity of Visual Memory
Abstract

Visual long-term memory can store thousands of objects with surprising visual detail, but just how detailed are these representations and how can we quantify this fidelity? Using the property of color as a case study, we estimated the precision of visual information in long-term memory, and compared this to the precision of the same information in working memory. Observers were shown real-world objects in a random color, and then ask to recall the color after a delay. We quantified two parameters of performance: the variability of internal representations of color (fidelity) and the probability of forgetting an object’s color altogether. Surprisingly, the data show that the fidelity of color information in long-term memory was comparable to the asymptotic precision of working memory. These results suggest that a common limit may constrain both long-term memory and working memory, such as a bound on the fidelity required to retrieve memory representations.
A large body of work has demonstrated that visual long-term memory is capable of storing thousands of objects with significant detail (Brady, Konkle, Alvarez, Oliva, 2008; Konkle, Brady, Alvarez & Oliva, 2010a, 2010b; Hollingworth, 2004, 2005). However, long-term memory fidelity has been examined in only a qualitative way. For example, Brady et al. (2008) demonstrated that, after seeing thousands of objects, observers succeeded at subtle object-exemplar discriminations (e.g., which of two chocolate cakes were seen) or object-state discriminations (whether the cake was half eaten or 2/3rds eaten). But the information observers must store to recognize the amount of cake is difficult to quantify and compare across timescales and items. So several fundamental questions remain unanswered: Just how detailed is visual long-term memory? And how does this compare to the detail contained in visual working-memory or the precision of perception?

Determining the precision of long-term memory places significant constraints on models of memory in general, and is particularly relevant for understanding the relationship between working memory and long-term memory. For example, if working memory and long-term memory have similar fidelity, then it is important to consider unified explanations for the limit on fidelity (e.g., fidelity-dependent retrieval limits), as opposed to system-specific limitations (e.g., the amount of "slots" or "resources" available to working memory; Zhang & Luck, 2008; Wilken & Ma, 2004). Thus, comparing working memory and long-term memory can not only help elucidate the underlying memory representation of visual objects (Brady, Konkle & Alvarez, 2011), but also clarify the extent to which these two stores rely upon shared representations and processes (McElree, 2006; Nairne, 2002; Jonides et al. 2008).
Previous attempts to quantify the fidelity of long-term memory representations used simple stimuli like oriented gratings (Magnussen & Dyrnes, 1994; Magnussen et al. 2003). However, this work may depend on memory for decision criteria rather than perceptual features of the objects-to-be-remembered (Lages & Treisman, 1998; Lages & Paul, 2006; Magnussen, 2009). In addition, while oriented gratings define a well-characterized space within which to quantify fidelity, these stimuli are not suited to the strengths of long-term memory, which is best studied using meaningful stimuli like real-world objects (Konkle et al., 2010). Thus little is known about how detailed visual long-term memory representations can be with real-world, semantically-rich objects.

Here, we take a psychophysical approach to quantify the fidelity of visual long-term memory for objects. We use color as a case study because the color of objects can be manipulated in a continuous space, which allows us to extend the continuous report paradigm used in visual working memory (Wilken & Ma, 2004) to long-term memory. Furthermore, there are metrics that allow separable measurements of the fidelity of color memory independent of guessing (Zhang & Luck, 2008; Bays, Catalao & Husain, 2009). Finally, previous work has shown that continuous report metrics do not seem to depend on verbal memory (Zhang & Luck, 2008) and that results with color generalize to shape (Zhang & Luck, 2008) and orientation (Anderson, Vogel & Awh, 2011). Thus, we can quantify how accurately observers remember the features of a given object after seeing hundreds of objects, and how likely they are to completely fail to retrieve a feature.

We find that from perception to working memory, observers lose significant precision in their representation of objects’ color. As more items are added to working memory, the fidelity of these memory representations reaches an asymptotic limit, and
surprisingly, this limit is almost identical to the fidelity of representations we find in long-term memory. These results suggest that a common limit may be at work in both visual working memory and long-term memory: the asymptotic fidelity observed in visual working memory may not be a consequence of a slot-like architecture (Zhang & Luck, 2008, 2009; Anderson, Vogel & Awh, 2011) or a limited pool of resources (Wilken & Ma, 2004; Bayes, Catalao, & Husain, 2009); rather, the fidelity of visual working memory may reflect a more general upper bound on how noisy a memory representation can be before it is unable to be retrieved.

**EXPERIMENT 1A AND 1B**

Observers performed a continuous color report task on pictures of real-world objects. Observers were shown an object with a randomly chosen hue and, after a delay, asked to choose from a color wheel what hue the object was. Such continuous report methods have been used with simple geometric shapes in working memory (e.g., Wilken & Ma, 2004; Zhang & Luck, 2008; Brady & Alvarez, 2011), but have never been adapted for long-term memory.

Importantly, this method allows us to measure the fidelity of perception, working memory and long-term memory using a within-subject design. In the *perception* condition, observers had to match the color of a visible object. In the *working memory* condition, observers were given 3 seconds to encode 3 objects, and then had to report the colors of each object after a 1 second delay. We used three objects to match the set size at which working memory fidelity asymptotes (Zhang & Luck, 2008; Anderson et al. 2011). In the *long-term memory* condition, observers viewed hundreds of objects, presented one at a time, and then were asked to report the color of every single object, one at a time. In
Experiment 1A, observers had 3 seconds to encode each object for the long-term memory condition, matching the total time that a display was visible in the working memory condition. In Experiment 1B we gave observers only 1 second to encode each object in the long-term memory condition, matching the time per-object in the working memory condition. Observers saw different objects in each of the three conditions.

**Methods**

*Participants*

14 observers (age range 18-25) participated in Experiment 1 -- five in Experiment 1A and nine in 1B. They gave informed consent, and had normal color vision (assessed using Ishihara’s test for color deficiencies). All participants completed each of the three conditions, with the order randomized across participants.

*Apparatus*

Experiments were run in MATLAB using the Psychophysics Toolbox (Brainard, 1997; Pelli, 1997), with stimuli presented on a 49° by 31° display, viewed from 57 cm.

*Stimuli*

540 pictures of categorically distinct objects were selected from Brady et al. (2008). The objects were chosen such that they consisted largely of a single arbitrary color (e.g., the object would be recognizable in any color; see Figure 1). Objects were rotated randomly in hue space, such that on each trial, the initial object color was determined by adding a
random angle between 0 and 359 to the original hue. Each image appeared equally often in the three conditions. All stimuli subtended ~6° of visual angle.

Figure 1: Methods of Experiment 1. Observers were shown objects with their colors randomly rotated in hue, and asked to report the object’s color, either while it was still visible ([A] Perception), after a 1s delay ([B] Working Memory) or after seeing several hundred objects over 30 minutes ([C] Long-Term Memory). In Experiment 1A, the items in long-term memory were shown for 3s/item, in Experiment 1B for 1s/item.

Perception Condition

To assess the fidelity of color perception, participants performed a color-matching task (see Figure 1). On each of the 180 trials, two copies of the same image were presented
simultaneously, centered 5° left and right of fixation. The left image was the “standard”
image, the right was the “test” image (initially in grayscale). The task was to adjust the
color of the test item to match the standard.

*Working Memory Condition*

On each trial, three objects were simultaneously presented for 3s, in a circle around
fixation. Participants were instructed to remember the color of all 3 objects. The objects
disappeared for 1s, and then memory for the color of each item was tested one at a time
in a randomly chosen sequence. Participants completed 60 trials, for a total of 180 tests.

*Long-Term Memory Condition*

*Study Block.* Participants viewed 232 images presented one at a time, for either 1s (1A) or
for 3s (1B) each. There was a 1s blank between images. Participants were instructed to
remember the color and identity of each object as they viewed the images. During the
study block, participants performed a repeat detection task to encourage them to maintain
focused attention. Twenty-six images in the study stream appeared twice in a row, and
participants responded by pushing the spacebar for the repeat. Participants were given
feedback only when they responded; with a red fixation cross indicating an incorrect
response or a green fixation cross indicating a correctly detected repeat. No feedback was
given for misses or correct rejections.
Test Block. Immediately after the study block, we tested the fidelity with which participants remembered the color of the objects. Items that were repeated in the study stream were not tested, leaving 180 tested images.

Continuous Report

For each of the three conditions, participants’ color memory was measured using the method of adjustment. At the beginning of each test, the item appeared in grayscale and the mouse pointer appeared at the center of the item. When the participant moved the mouse, the test item appeared in color. The angle between the mouse and the center of the test item determined the item’s hue, and a dot was presented along the adjustment ring indicating the current position. When participants decided that the current color was correct, they clicked the mouse. The angular error was taken as a measure of the accuracy. The color wheel was randomly rotated across trials.

Participants proceeded at their own pace and were asked to be as accurate as possible in their decisions. Feedback was given after accurate responses. The words “good”, “great” or “perfect” appeared on the screen for errors of less than 10°, 5° or 0°.

Data Analysis

On any given trial we measured error in degrees, between 0 (perfect memory) and +/- 180 degrees (poor memory). The histogram of errors over trials shows that most responses are centered around 0, but that across all responses there are errors distributed across the entire range (Figure 2). This error histogram is well-fit by a mixture of two distributions: (1) a Gaussian-like distribution (defined on a circular space as a von Mises
distribution), taken to reflect successful memory retrieval with some degree of precision, and (2) a uniform distribution, taken to reflect random guessing trials. Following Zhang and Luck (2008), this method allows us to separate trials in which the color was retrieved with some level of fidelity, and trials in which the color of the item was forgotten.

The fidelity of memory representation was estimated as the standard deviation (SD) of the von Mises distribution. The narrower the distribution appears around 0, the more precise the memory representation. The probability of guessing (PG) is estimated by the height of the uniform distribution. Maximum likelihood estimation was used to estimate these two parameters for each condition.

**Results**

*Experiment 1A*

In the perception condition, observers were highly accurate, with a precision estimated at SD = 6.7° (+/- 0.8 S.E.M.) and the probability of guessing estimated at PG = 0.0 (+/- 0.0 S.E.M.). Thus, when the stimulus was present on the screen, observers’ never responded randomly and had a tight distribution centered on the correct color.

In the working memory condition, in line with Zhang & Luck (2008), observers’ SD was 19.0° (+/- 1.3 S.E.M.) and guess rate PG was 0.09 (+/- 0.02 S.E.M.). This represents a major change in fidelity from perception to working memory: the SD increases by 183%, a serious cost in memory fidelity for having to hold the items in mind for several seconds (difference: \( t(4)=10.5; p<0.0001 \)).

In the long-term memory condition (with 3s/item), observers SD was 20.3° (+/- 3.3 S.E.M.), with a guess rate, PG, of 0.58 (+/- 0.05 S.E.M.). The increase in guess rate
from working memory was quite large, from 0.09 to 0.58 \( (t(4)=17.5, p<0.0001) \).

However, surprisingly, the fidelity (SD) observed for 180 items in long-term memory was not significantly different from the fidelity (SD) observed for three items at a time in working memory \( (t(4)=-0.81; p=0.46) \). Note that the precision of working memory estimated here is similar to the precision observed across several working memory experiments testing memory for color patches at set sizes 3 and greater (Zhang & Luck, 2008; Zhang & Luck, 2011) despite our use of real-world objects.

**Experiment 1B**

Observers had only 1s to encode the color of the item in long-term memory condition. Despite this severe decrease in encoding time, Experiment 1B replicated Experiment 1A nearly exactly (see Figure 3). The fidelity (SD) in perception was 4.7° (+/- 0.5 S.E.M.), working memory was 17.8° (+/- 1.0 S.E.M.), long-term memory was 19.3° (+/- 0.9 S.E.M.). The probabilities of guessing (PG) were 0.006 (+/- 0.002), 0.08 (+/- 0.01) and 0.63 (+/- 0.05) respectively. As before, the fidelity of working memory and long-term memory were not significantly different \( (t(8)=1.0, p=0.33) \). In addition, there was no significant difference between the fidelity of long-term memory across the two experiments \( (t(12)=0.36, p=0.72) \). Thus, the extra encoding time made no difference to the fidelity of color information in long-term memory. Combining results across Experiments 1A and 1B also did not result in a significant difference between the fidelity of working memory and long-term memory \( (means: 18.2, 19.7, t(13)=0.84, p=0.41) \).

Figure 2 shows the distribution of errors for the perception, working memory, and long-term memory conditions, combining across Experiment 1A and Experiment 1B.
The summary of the fidelity and guessing parameters for each experiment are presented in Figure 3.

**Figure 2**: Results pooled across Experiments 1A and 1B. The histograms represent the error in observers’ responses for (A) Perception, (B) Working Memory, and (C) Long-term Memory (pooled across all observers). While the fidelity of color representations decreases between perception and working memory, nearly all of the difference between working memory and long-term memory is driven by an increased guess rate.
Figure 3: Summary of results from Experiments 1A and 1B. While guess rate goes up significantly between working memory and long-term memory, the fidelity of long-term memory is comparable to that of working memory -- in spite of the need to represent many more items for a much longer duration, and the need for observers to encode and retrieve the items rather than keep them actively in mind.

Discussion

We measured the fidelity of color information in visual long-term memory in two studies, and compared it to the fidelity of working memory and perception. The data show an extremely precise fidelity in perception (~5 degrees in hue), which falls off significantly in both working memory and long-term memory. Surprisingly, the fidelity for color was comparable between working memory and long-term memory (standard deviation of ~20 degrees). This was true when long-term memory encoding times matched the total time (Experiment 1A) or the per-item time of the working memory.
condition (Experiment 1B). The results showed that nearly all of the information loss from working memory to long-term memory is accounted for by an increased chance of entirely losing an item’s color from memory (increased guess rate).

In the long-term memory condition, people must store hundreds of items for long durations and are required to encode and then retrieve them; whereas in the working memory condition, observers can keep items and their colors actively in mind. Despite these major differences between the two tasks, when holding 3 items in working memory, the fidelity of working memory and long-term memory are nearly identical. This indicates that observers have highly detailed long-term memory representations – even when using quantitative measures rather than qualitative forced-choice comparisons (Brady et al. 2008, Hollingworth, 2004).

**EXPERIMENT 2**

It is possible that long-term memory and working memory have the same fidelity because long-term memory representations directly inherit their fidelity from working memory. For example, if items have to enter working memory to be encoded into long-term memory, and if there is no further degradation of representations once encoded, then long-term memory representations would have exactly the same fidelity as working memory representations. While this is a possible account of our results, previous work has shown that the fidelity of visual working memory depends on the number of items remembered (Zhang & Luck, 2008; Wilken & Ma, 2004). Thus, because items in our long-term memory task were presented sequentially, one at a time, this inherited precision account would predict that the precision of long-term memory representations
in Experiment 1 should match the fidelity of working memory for a single item. Experiment 1 cannot directly address this prediction, because multiple items were presented simultaneously in the working memory task.

To address this *inherited precision* hypothesis, we matched the encoding conditions in the working memory and long-term memory conditions in Experiment 2. To preview the results, we found that observers can in fact remember a single item in working memory with better precision than long-term memory, which rejects the inherited precision account.

**Methods**

**Participants**

Six observers participated in Experiment 2. None had previously participated in Experiment 1. All participants gave informed consent, were between the ages of 18 and 25, and had normal color vision (assessed using Ishihara’s test for color deficiencies).

**Stimuli**

The stimuli were the same objects as in Experiment 1.

**Procedure**

The procedure was identical to Experiment 1A, with the exception of the working memory condition, which was modified to consist of 180 trials with only a single item presented on each trial. Each object was presented for 3s and then tested after a 1s delay.
Results

The perception (SD = 5.6° +/- 1.1, PG = 0.01 +/- 0.01) and long-term memory conditions (SD = 20.5° +/- 7.0, PG = 0.67 +/- 0.15) replicated the results from Experiments 1A and 1B. However, working memory fidelity for 1 item was significantly better than long-term memory fidelity, given matched encoding conditions (SD=14.5° +/- 1.3; t(5)=2.96, p=0.03). In addition, comparing across experiments, the fidelity of working memory for 1 item was significantly better than the fidelity for 3 items (t(9)=2.45, p=0.03 and 1B: t(13)=2.06, p=0.06). These results show that the precision of color information of real-world object stimuli is not fixed at encoding; it is possible for the fidelity of working memory to be better than the fidelity of long-term memory.

Discussion

The results of Experiment 2 show that when a single real-world object is encoded, it has a higher fidelity representation in working memory than it will when it is later probed in long-term memory. This indicates that the fidelity of long-term memory is not directly inherited from working memory, and that there is additional degradation in long-term memory that reduces the measured precision of retrieved items to 20 degrees.
Figure 4. The standard deviation (measure of fidelity) from all experiments plotted on a single axis. Perceptual fidelity is plotted for three different experiments (Experiments 1A, 1B, and 2). Working memory fidelity is plotted as a function of the number of items on the display, for 1-item displays (Experiment 2), 3-item displays (Experiment 1A and Experiment 1B), and a 5-item display (Control Experiment 1). Long-term memory precision is plotted for 6 different experiments (Experiments 1A, 1B, 2; a control experiment with 360 items rather than 180; a control experiment with only 20 items; and a control experiment with verbal interference; see Supplemental Online Material). In line with previous work, we find that as you add more items to working memory, fidelity plateaus. Interestingly, the fidelity at which working memory reaches a plateau is the same as the fidelity estimated for visual long-term memory (highlighted with a gray zone). The solid line for working memory shows the fidelity as a function of set size. The dashed lines for perception and long-term memory indicate the mean standard deviation for those conditions.

Intriguingly, the data across all experiments show that the level at which working memory plateaus is identical to the fidelity observed in long-term memory (Figure 4). To bolster this finding, we conducted several control experiments (see also Supplemental Online Material).
First, we tested working memory using displays with 5 items, to more clearly demonstrate the plateau in the fidelity of working memory over these 1-, 3-, and 5-item displays (Experiment C1, illustrated in Figure 4, green line). Next, we asked whether we could make long-term memory precision worse than this limit. We reasoned that if we doubled the number of items in memory from 180 to 360, this might lead to less precise memories. However, we instead found that this manipulation only increased the probability of guessing, and the fidelity of the remembered item remained at about SD=20 degrees (Experiment C2). Finally, we examined whether long-term memory precision could be more precise than this limit. We note that if long-term memory contained more precise memory representations, this would still be consistent with a limit on the fidelity of the memory representations. However, surprisingly we found that even with only 20 items in mind, precision was similar, and the benefit in overall performance was reflected with lower guessing rates than in the case of 180 or 360 items (Experiment C3). This result was also obtained when participants performed a verbal interference task (Experiment C4), suggesting this limit is unlikely due to a verbal coding strategy.

In summary, across a very wide range of overall difficulty levels in the long-term memory task, where guess rates ranged from 26% to 73%, we find that fidelity remains constant at a standard deviation of 19-20 degrees. Any item more variable than this limit seems to be lost entirely to guessing, in both working memory and long-term memory.

**GENERAL DISCUSSION**

Across several experiments, we find that observers lose significant precision in their representation of real-world objects when going from perception to working
memory. However, the precision of three or four actively maintained representations in working memory is the same as that of hundreds of representations encoded and then retrieved from long-term memory. Thus long-term memory fidelity is significantly higher than previous believed, even when quantified using psychophysical methods.

Furthermore, the fidelity of long-term memory is not directly inherited from the fidelity of representations at encoding, but instead seems to represent an asymptotic limit on the fidelity of items retrieved from memory: In working memory, as the number of items stored increases, fidelity plateaus with a standard deviation of about ~20 degrees. Similarly, when retrieving items from long-term memory, fidelity appears to degrade, but for remembered items fidelity does not get worse than a standard deviation of approximately 20 degrees, despite the necessity of representing more items for a longer duration and making use of an encoding and retrieval process rather than active storage. Additionally, this degree of fidelity is robust to a variety of long-term memory encoding durations, the number of objects stored, and the presence or absence of a verbal interference task. Thus, we suggest that a standard deviation of ~20 degrees may represent a limit on the fidelity of arbitrary color information that can successfully be retrieved from memory: any memory representations that degrade so that they have more variability than an SD of ~20 degrees seem to be irretrievable.

This pattern of results suggests a dramatic reinterpretation of existing data from working memory: the plateau in working memory fidelity is likely not caused by factors intrinsic to working memory like the fidelity of a slot (Zhang & Luck, 2008, 2009; Anderson, Vogel & Awh, 2011) or the quantity of a resource (Bays et al. 2009), but is instead a general property of the memory encoding and retrieval system. That is, the
fidelity of working memory and long-term memory may reflect an upper bound on how noisy a memory representation can be before it is unable to be retrieved.

**Figure 5.** Pictorial representation of the memory fidelity we observe in perception, working memory (3 items) and long-term memory. The central stimuli are the studied color, and the items next to them represent +/- 1 standard deviation (6 degrees in perception, 19 degrees in working memory for 3 items, 20 degrees in long-term memory).

**Relationship between working memory and long-term memory**

Several influential studies have found that working memory fidelity plateaus at a standard deviation of approximately 20 degrees (Zhang & Luck, 2008, 2009, 2011; Anderson et al. 2011; Anderson & Awh, 2012). In particular, fidelity does not seem to decrease when more than 3 or 4 items are encoded (Zhang & Luck, 2008), or when observers hold items for longer durations (Zhang & Luck, 2009). Based on this apparent asymptote in fidelity, these studies have concluded that working memory represents items with discrete slots that undergo catastrophic failures when items are held for long durations (Zhang & Luck, 2009). However, these explanations for why fidelity does not
become worse are based entirely on models of active storage in working memory (slots, resources). For example, Zhang and Luck (2008) interpreted this asymptote as resulting from a limited number of memory slots, which maintain fixed-resolution representations in working memory. According to this theory, if you have 3 slots in memory, you could use them to represent < 3 items with more precision by allocating multiple slots per item, but after you had 3 items in memory you could no longer split your representations among more items, and so all subsequent items were simply not encoded. This failure to encode more than 3 items results in an increased guess probability but a flat fidelity asymptote as the number of items exceeds the number of slots. Other researchers have argued that the asymptote is a natural consequence of spreading a continuously divisible memory resource across multiple items, leading to decreased precision and an increased likelihood of forgetting items as set size increases (e.g., Wilken & Ma, 2004; Bays & Husain, 2008).

However, our finding that long-term memory shares similar limits suggests an alternative model. Rather than this asymptotic fidelity limit reflecting an intrinsic property of the working-memory system, it may instead reflect a property of the broader memory system and factors that limit memory retrieval.

**Conclusion: Shared limits for an integrated visual memory system**

The broader working memory literature—particularly for verbal stimuli —has accumulated significant evidence for shared principles between short-term and long-term memory (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). In addition, a number of empirical results highlight that
working memory tasks do not solely isolate working memory mechanisms independent
of long-term mechanisms. For example, performance on any given working memory trial
is influenced by previous trials, including systematically induced biases and proactive
interference (Huang & Sekuler, 2010; Hartshorne, 2008; Makovski & Jiang, 2008),
suggesting an obligatory influence of long-term storage on working memory (Brady,
Konkle, Alvarez, 2011; see also Olson et al. 2006, for evidence from neuroscience).

The present empirical results showing that long-term memory fidelity is so high,
and, in fact, equivalent to the asymptotic fidelity of working memory, leads us to propose
a new link between working memory and long-term memory: both appear to have the
same lower bound on memory fidelity. Recalled items never appear noisier than a fixed
limit; after that limit, they are lost, perhaps unable to be retrieved for conscious access
because they no longer sufficiently resemble the original memory trace that was laid
down.

ACKNOWLEDGEMENTS

Stimuli may be found on the authors’ website, and http://cvcl.mit.edu/MM/. Work in part
supported by the National Science Foundation under Grant No. 1016862 and a faculty
research award from Google to A.O.

REFERENCES

Reaches a Stable Plateau When Individual Item Limits Are Exceeded. Journal of
Neuroscience, 31(3), 1128-1138.


