Journal of Experimental Psychology: General
2017, Vol. 146, No. 11, 1551–1573

0096-3445/17/$12.00 http://dx.doi.org/10.1037/xge0000346

The Nature of Short-Term Consolidation in Visual Working Memory

Timothy J. Ricker
College of Staten Island and The Graduate Center, City University of New York

Kyle O. Hardman
University of Missouri

Short-term consolidation is the process by which stable working memory representations are created. This process is fundamental to cognition yet poorly understood. The present work examines short-term consolidation using a Bayesian hierarchical model of visual working memory recall to determine the underlying processes at work. Our results show that consolidation functions largely through changing the proportion of memory items successfully maintained until test. Although there was some evidence that consolidation affects representational precision, this change was modest and could not account for the bulk of the consolidation effect on memory performance. The time course of the consolidation function and selective influence of consolidation on specific serial positions strongly indicates that short-term consolidation induces an attentional blink. The blink leads to deficits in memory for the immediately following item when time pressure is introduced. Temporal distinctiveness accounts of the consolidation process are tested and ruled out.

Keywords: attentional blink, consolidation, short-term memory, visual memory, working memory

When stimuli are perceived sensory memory traces are created, providing a representation that is accessible to conscious awareness. These representations are lost quickly if they aren't attended and processed further (Cowan, 1984; Massaro, 1975; Sperling, 1960). Working memory contrasts with sensory memory in that it is much more limited in the amount that can be represented, but the representations that are maintained are more durable and some subset of them can be maintained as long as active maintenance activities are carried out (Barrouillet & Camos, 2012; Cowan, 1988; Oberauer, Lewandowsky, Farrell, Jarrold, & Greaves, 2012). The memory traces maintained in working memory are available for use in any ongoing cognitive task without the need for retrieval from long-term storage, providing a mental workspace for cognition (Baddeley, 2000; Cowan, 1995). Here we examine the process of transforming transient sensory memory traces into more durable working memory traces. This process is generally known as short-term consolidation (Jolicœur & Dell'Acqua, 1998; Ricker, 2015; Ricker & Cowan, 2014).

Previous work has shown that short-term consolidation requires attention but can be completed even when the consolidated item is no longer physically present in the perceptual field. Short-term consolidation can occur as long as attention is allowed to dwell on the internal memory representation (Jolicœur & Dell'Acqua, 1998; Ricker, 2015; Ricker & Cowan, 2014). Although it requires attention, short-term consolidation is completed relatively quickly with estimates of its time-course ranging from 400 to 1600 ms (Wyble, Potter, Bowman, & Nieuwenstein, 2011; Nieuwenstein, Potter, & Theeuwes, 2009; Stevanovski & Jolicœur, 2007). During this period, execution of other attention-demanding tasks does not proceed (Jolicœur & Dell'Acqua, 1998; Nieuwenstein & Wyble, 2014; Ouimet & Jolicoeur, 2007; Stevanovski & Jolicœur, 2007). Beyond this very rough time-course, not much else is known about what occurs during the consolidation process.

Our knowledge of short-term consolidation is largely defined by evidence for what the process is not, rather than by evidence of what the process involves. Some investigations into the consolidation process initially seemed to show that it is stopped by perceptual masking (Vogel, Woodman, & Luck, 2006; Woodman & Vogel, 2005, 2008). Although these studies do show that masking a stimulus immediately after its presentation does lead to performance deficits, recent evidence demonstrates that short-term consolidation clearly continues after masking stimuli are presented (Jolicœur & Dell'Acqua, 1998; Nieuwenstein & Wyble, 2014; Ricker & Cowan, 2014). Other evidence shows that consolidation effects are not related to initiation of articulatory rehearsal or related verbalization processes as preventing rehearsal does not alter the consolidation effect (Bayliss, Bogdanovs, & Jarrold, 2015; Stevanovski & Jolicœur, 2007). Bayliss et al. also provide evidence that suggests consolidation and attention-based maintenance are fundamentally different processes.

Although limited in scope, we do have some knowledge about what occurs during the consolidation process. We have strong evidence that the consolidation process relies on attention as other attention-demanding tasks are delayed while memory stimuli are

This article was published Online First July 13, 2017.

Timothy J. Ricker, Department of Psychology, College of Staten Island, and The Graduate Center, City University of New York; Kyle O. Hardman, Department of Psychological Sciences, University of Missouri.

Some of the data and ideas in this article were presented at the 2015 and 2016 Meetings of the Psychonomic Society. Some of the work reported in manuscript was supported by NIH Grant R01-HD21338 to Nelson Cowan. Thanks to Jeff Rouder for allowing us to run two of our experiments in his laboratory facilities. Thanks to April Swagman, Zach Shipstead, and Brad Wyble for their helpful comments on earlier versions of this article.

Correspondence concerning this article should be addressed to Timothy J. Ricker, Department of Psychology, College of Staten Island, 2800 Victory Blvd, 4S-234, Staten Island, NY 65211. E-mail: timothy.ricker@csi.cuny.edu

consolidated (Jolicœur & Dell'Acqua, 1998; Nieuwenstein & Wyble, 2014; Ouimet & Jolicoeur, 2007; Stevanovski & Jolicœur, 2007). We also have evidence that short-term consolidation improves working memory performance by reducing forgetting (Ricker & Cowan, 2014). Evidence that consolidation functions by reducing later forgetting supports a theory of the consolidation process as stabilization of memory traces that are initially quite fragile (Ricker, 2015). Consolidation reduces future forgetting by increasing representation robustness.

Most previous work on short-term consolidation in the working memory literature has inferred that a consolidation process exists without actually observing changes to memory performance. Instead changes in secondary task performance have been used to infer that short-term consolidation is operating in a concurrent memory task. Generally memory items are presented and a secondary task, such as tone identification, is performed within the next few seconds. Delays or reduced performance in the secondary task are observed when the temporal proximity between the two tasks is particularly close. Only recently have working memory tasks been directly shown to suffer reduced performance when short-term consolidation times are limited (Nieuwenstein & Wyble, 2014; Ricker & Cowan, 2014). This is a crucial step in confirming that changes to the memory traces are occurring during the consolidation period. Here we take the next step and seek to understand how the maintained memories benefit from short-term consolidation.

We explore several potential mechanisms through which short-term consolidation could improve working memory performance. We start by explaining these theoretical approaches to understanding performance benefits from short-term consolidation based on the existing literature. Following this, we derive clear predictions from each theoretical approach to specify the actual mechanism(s) by which short-term consolidation improves working memory performance. We then test these predictions across 4 experiments and discuss what the results tell us about the nature of short-term consolidation.

Potential Mechanisms Causing Short-Term Consolidation Benefits in Working Memory

Stabilization of Representational Precision

A simple theory of short-term consolidation is that the consolidation process leads to a change in the precision of stimulus representation that affects performance at the time of test. Time may be required to stabilize the details of the memory representation, with more consolidation time resulting in slower loss of memory precision during memory retention. A study by Bays, Gorgoraptis, Wee, Marshall, and Husain (2011) would seem to favor this view, finding that increasing presentation time increases the representational precision of visual items. However, manipulation of presentation time is different from manipulating the time available for consolidation. Consolidation continues to occur after an item is removed from the perceptual field and after perceptual masking. The Bays et al. study confounds presentation time and consolidation time, although this was not a problem for their goals as they were not studying consolidation.

A Change in the Nature of the Stabilized Representation

Often the general gist of a short-term representation is sufficient to perform the task at hand, making detailed stimulus representation unnecessary. When creating a memory from a perceptual trace we may initially stabilize only a very coarse categorical representation of the memory item and next stabilize the specific fine-details into a durable representation as a second step when we have enough time. For example, we may first consolidate a color as categorically red but then, as a second step, consolidate the exact shade of red presented. This differs from the preceding explanation of consolidation as stabilization of representational precision in that this approach does not propose a gradual change in a continuous dimension. Instead this approach suggests that consolidation occurs in multiple steps, first consolidating a categorical memory then consolidating a qualitatively different fine-detailed memory. If only a categorical memory is consolidated for a given stimulus then the fine-details of the stimulus, such as the shade of red, should be quickly forgotten leaving only the categorical information, the general category red, at test.

Recent work has provided support for multiple qualitatively different sources of information stored in working memory, one categorical and one with fine-detail. Bae, Olkkonen, Allred, and Flombaum (2015) show that participants incorporate both the categorical color and the precise shade of color presented when asked to reproduce the hue of a studied color at test. Hardman, Vergauwe, and Ricker, (2017) demonstrate that these two sources of information, categorical and continuous-detail for color, are represented as distinct memory states and only one is used to make a recall response on any given trial. Donkin, Nosofsky, Gold, and Shiffrin (2015) also asked participants to reproduce the hue of studied colors at test. They found that mathematical models of participant recall fit the data better when they incorporated memory for verbal color labels in addition to memory for the specific color presented. These verbal labels could be another way of thinking of coarse memory states that are separable from precise visual representations of stimuli.

Given the evidence for multiple memory states in working memory tasks it is reasonable to assume that these differing states would need to be consolidated separately. The simpler categorical representation would be a quick and easy initial target for consolidation while a second more-detailed representation would likely take longer to stabilize and provide a more detailed representation of the stimulus.

Avoiding an Attentional Blink

Evidence from research on the attentional blink and related paradigms can lead to a different approach to thinking about the benefits gained from working memory consolidation. The attentional blink refers a brief period of time, roughly 200–700 ms, after attentional capture by a target stimulus during which attention cannot be reallocated to new stimuli that appear (Chun & Potter, 1995; Raymond, Shapiro, & Arnell, 1992). This occurs when the identity of the target stimuli must be reported later and effectively renders the participant unable to identify new stimuli during the blink. Recent theories of the attentional blink have proposed that the blink is caused by consolidation of the initial stimulus identity into working memory for later report (Lagroix, Spalek, Wyble, Jannati, & Di Lollo, 2012; Wyble, Bowman, & Nieuwenstein, 2009; Wyble et al., 2011). According to these theories, consolidation into working memory suppresses attention preventing the use of attention for other purposes.

In the context of working memory performance, one can imagine a sequence of stimulus presentations such as that in Figure 1. The sequential presentation of stimuli is fairly common within the working memory literature and even when stimuli are presented concurrently they may be consolidated in serial, producing covert serial processing. When items are presented in close temporal proximity or when simultaneous array presentations are presented for short periods, failure to consolidate all of the presented items may occur because the attentional blink suppresses the ability to attend to all of the items within the time provided. Items that are not consolidated, or are only partially consolidated, should experience greater rates of forgetting because their perceptual representations have not been stabilized (Ricker, 2015).

A Change in Temporal Distinctiveness

Temporal distinctiveness theories claim that all memory must be retrieved through long-term storage mechanisms at the time of recall (Brown, Neath, & Chater, 2007; Crowder, 1976; Souza & Oberauer, 2015; Shipstead & Engle, 2013). In these theories memory retrieval success is determined by the distinctiveness of the item being retrieved. This distinctiveness is determined by how different the target memory is from other similar memories along a number of psychological dimensions, including a temporal one. Increasing the time available for consolidation of each item will also increase the temporal spacing between memory items. This will increase the temporal distance between memory items thereby boosting memory discriminability. From this view, increases in performance following longer consolidation periods are actually attributable to increased temporal discriminability and not to a consolidation process.

The Present Approach and Predictions Following From Each Potential Mechanism

In the four experiments that follow we present four visual items in sequential order and vary the time between item presentations to vary the amount of time available for consolidation (see Figure 1). Our method builds on recent work in which the investigation of visual working memory has shifted toward an emphasis on recall rather than simply recognition or change detection (Fougnie, Suchow, & Alvarez, 2012; Hardman et al., 2017; Ma, Husain, & Bays, 2014; Zhang & Luck, 2008). In a delayed estimation task (Wilken & Ma, 2004; Zhang & Luck, 2008), participants are asked to reproduce the observed memory stimulus. Because visual features such as color and orientation vary in a continuous manner, a continuous response error can be observed in these tasks rather than simpler binary correct/incorrect responses. Mathematical models of the data can then be applied which specify the precision of the maintained representation (van den Berg, Awh, & Ma, 2014; Zhang & Luck, 2008) as well as the mixture of precise continuous information and gross categorical information (Bae et al., 2015; Donkin et al., 2015; Hardman et al., 2017).

The Present Model-Based Approach

In the present experiment we examine consolidation effects within visual working memory using the delayed estimation technique and a model of performance developed by Hardman et al. (2017) that accounts for both precise representation of the specific angle presented and categorical knowledge of the general direction of the presented angle. Previous work on consolidation effects has

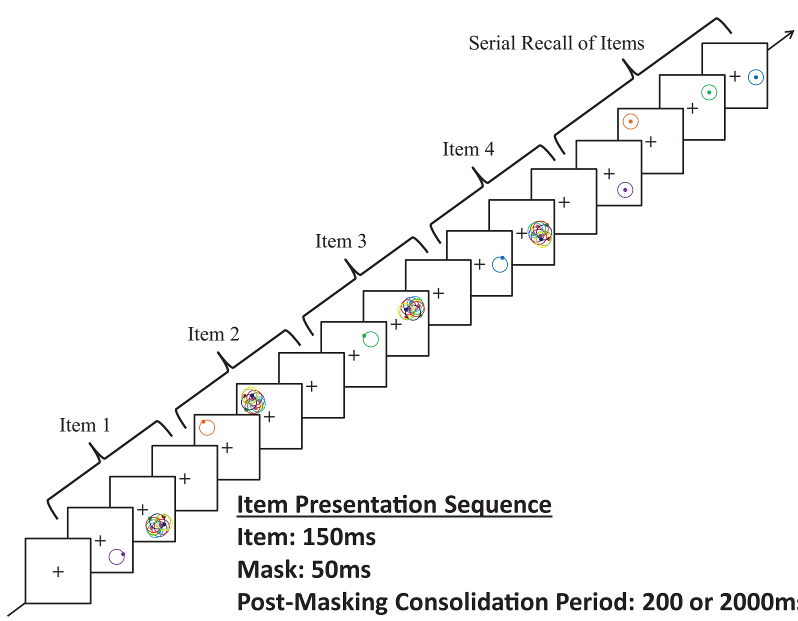

Figure 1. An example of a single experimental trial in Experiment 1. See the online article for the color version of this figure.

used recall of verbal materials or recognition of visual materials. We present visual stimuli which vary in angular orientation and ask participants to reproduce the angle of the memory item at test (see Figure 1). The stimuli and responses vary continuously, giving us detail about the nature of the representation that is not available from binary correct or incorrect accuracies. Through this additional detail we can observe the nature of the changing representation during visual working memory consolidation, be it an increase in representational precision, a change in the probability of being successfully remembered, or a change in the nature of the representation itself between categorical memory and precise continuous memory.

Predictions Following From Each Theoretical Mechanism

The four theoretical mechanisms that we explain above each make differing predictions about how model parameters should change as a function of consolidation time. By varying the consolidation time allowed and applying the mathematical model of Hardman et al. (2017) to the resulting data we can identify which of these mechanisms are responsible for the short-term consolidation effect in working memory.

Our first theoretical mechanism, *stabilization of representational precision*, should result in changes to the standard deviation of fine-detailed specific-angle responses. Stabilization of the representation during consolidation results in a representation that has improved precision throughout maintenance and at test. This approach predicts that specific-angle responses should have a smaller standard deviation when more consolidation time is given.

Our second theoretical mechanism, *a change in the nature of the stabilized representation*, should result in a change in the proportion of responses that contain specific-angle information relative to only categorical angle information. With longer consolidation time more responses should contain stabilized representations of the exact angle presented at test. Within the model of Hardman et al. (2017) this translates to a change in the parameter representing the probability that a specific-angle response is produced, given that a memory is present.

Our third theoretical mechanism, *avoiding an attentional blink*, should result in a change in the probability that a memory of the stimulus is present at test (Asplund, Fougnie, Zughni, Martin, & Marois, 2014; but see, Sy, Marois, & Tong, 2015). Consolidation explanations of the attentional blink characterize the blink as occurring because a working memory state is being created (Wyble et al., 2009, 2011). This process is thought to fully occupy or suppress attentional resources, depending on the theoretical perspective. If consolidation time is too short one of two failures may occur, either of which leads to a failure to establish a stable working memory representation. The first failure happens if attention is reallocated too quickly, in which case the consolidation process will not have had time to be completed. This results in a failure to create a stable working memory trace of the first item. The second failure is possible when the second item onsets before consolidation of the first is complete and consolidation of the first item continues. Fully consolidating the first item would sometimes result in failure to shift attention to the newly presented stimulus while it is available in sensory memory and failure to consolidate the second stimulus. In either case the predicted result is a decrease in the probability that memory items are in memory at test under shorter consolidation periods, although which memory item is affected depends on the specific failure, which we examine in Experiment 4.

Our fourth theoretical mechanism, *a change in temporal distinctiveness*, should also result in a change in the probability that a memory of the stimulus is present at test. In this approach changes in consolidation time affect memory performance by modifying the retrieval discriminability at test (Brown et al., 2007; Shipstead & Engle, 2013). Failure to retrieve the proper memory item will result in responses that are not based on the stimulus presented. In the parameterization of the Hardman et al. (2017) model responses not based upon the memory stimulus are called guesses. An increase in the guess rate is equivalent to a decrease in the probability that an item is in memory at test. Shorter consolidation times should result in lower temporal discriminability and a resulting lower probability that the correct memory representation is retrieved for use at test.

Note that our third and fourth theoretical mechanisms make similar predictions. In particular, both predict that a decrease in consolidation time should result in fewer items being successfully recalled. The critical distinction between the predictions is the locus of effect. The attentional blink account predicts that items will be lost either before or after a short consolidation period, so that the effect would only be observed adjacent to a short consolidation interval. The temporal distinctiveness account would instead predict that a short consolidation interval within a list would affect the whole list, with the magnitude of the effect being largest nearest to the short consolidation interval and decreasing for items farther away from the short consolidation interval. We were able to distinguish between these two predictions in Experiment 4.

Experiment 1

Our first experiment is designed to produce a strong consolidation effect. As shown in Figure 1, we present four rings, each with a dot on the edge, one at a time and vary the amount of consolidation time allowed between item presentations while maintaining a constant presentation time across conditions. Consolidation time was manipulated between trials, with all consolidation periods on a single trial being equal in length. After all items are presented participants recall the location of each of the dots in the order that they were presented.

By modeling the distribution of responses produced at recall we can identify whether changes in consolidation time result in more precise memories, qualitatively different memory states, or a higher probability of remembering the memory item at test. These changes each map onto a different theoretical mechanism discussed above, with the exception that a change in the probability of remembering the memory item at test could indicate an effect of either an attentional blink or temporal distinctiveness.

Method

Participants. Thirty-four students (aged 18–20, 21 female) enrolled in introductory psychology at the University of Missouri participated in exchange for partial course credit. All participants had normal or corrected-to-normal vision.

The number of participants was determined by considering a mixture of design power and convenience. We first determined a minimum number of acceptable participants then collected data until that target was reached. At this point we continued collecting data from participants who had already scheduled experimental sessions but did not schedule new sessions. The minimum acceptable number of participants in Experiment 1, 24, was determined based upon past experience with similar paradigms in our lab. This number of participants both prevents instability in condition means and allows us to observe small effects in the data with the present experimental approach.

Materials. Each memory item consisted of a thin circular ring with a dot located at some point along the edge of the ring. The rings were 2.5 cm in diameter while the dot was 0.4 cm in diameter. The location of the dot was determined randomly with the constraint that it had to be within ±70 degrees in angle from the top of the circle. The memory set consisted of a series of four memory items, each presented individually. A masking stimulus was presented after each memory item. Masking stimuli were each composed of eight rings each slightly displaced from the location of the memory stimulus with eight dots presented in a random location near the location of the memory item. All eight masking rings and dots appeared concurrently. The color of each stimulus was determined by the location presented (see the procedure section for a description of the eight possible locations), such that each location was paired with a unique color for the entire experiment. For example, the stimulus directly to the right was always blue. Masking stimuli were composed of one ring and one dot of each of the eight colors. See Figure 1 for an example of the memory and masking stimuli.

Design. Participants were instructed to reproduce the location of the dot on the edge of each ring at test. There were two conditions with differing amounts of time available for consolidation. Condition was manipulated within participants. The short consolidation period was 200 ms in length, resulting in a total Consolidation Time of 400 ms from item onset, while the long consolidation period was 2000 ms in length, resulting in a total Consolidation Time of 2200 ms from item onset. Participants completed 8 practice trials followed by 4 blocks of 30 experimental trials. Each trial contained 4 delayed estimation responses resulting in 480 total experimental observations per participant.

Procedure. Figure 1 depicts a sample experimental trial. Participants initiated each trial by button press. At the start of each trial a fixation cross was presented for 500 ms. Next a memory item was presented on the screen for 150 ms in 1 of 8 possible locations. The possible locations for presentation of each memory item were directly above, below, to the right, and to the left of the center of the screen, as well as the four locations that would result from rotating this set 45 degrees. Memory items were always presented 4.5 cm from the center of the screen. Participants were instructed to remember the location of the dot on the edge of the ring for later recall. The memory item was immediately followed by the masking stimulus which was presented for 50 ms, then the consolidation period which was either 200 or 2000 ms in duration, determined randomly. Consolidation period duration was manipulated between trials, with all consolidation periods on a single trial being equal in length. This sequence, consisting of memory item—masking stimulus—consolidation period, was repeated four times in total with a different memory stimulus and location used in each item presentation sequence. Stimulus location was determined randomly from the eight possible locations. The same location was never used for multiple items on the same trial.

After all four items were presented, memory was tested. At test all four rings were presented, one at a time, at the same location and in the same order as at study. At test the dot was located in the center of the ring, not at the edge. Participants were instructed to move the dot to its original position by moving the mouse. When they were satisfied with the location of the dot they were to press a button on the mouse and the next ring would appear. After all items were tested, feedback was given. On the feedback screen all four rings were displayed, with the location of the participant's response in its original color and the correct response as a white dot slightly larger than the presented dot. One of three tones played depending on the mean error across items for that trial. A quickly rising series of tones played if mean error was 10 degrees or less. A slowly rising series of tones played if mean error was greater than 10 but within 30 degrees of the correct response. A falling series of tones played if mean error was more than 30 degrees.

Analysis. We report mean response error and conduct standard ANOVA testing to confirm any observed differences between conditions or across serial positions. We differ from the standard approach to inference and do not rely on interpreting *p* values to determine the presence or absence of a difference across conditions. Instead, we report the Bayes factor associated with our ANOVA results as this statistic has many advantages over *p* value-based inference (Goodman, 1999; Kass & Raftery, 1995; Wagenmakers, Wetzels, Borsboom, & van der Maas, 2011). The Bayes factor gives the probability of the observed data under one hypothesis relative to the probability of the data under an alternative hypothesis (Edwards, Lindman, & Savage, 1963; Rouder, Speckman, Sun, Morey, & Iverson, 2009; Wagenmakers, 2007). Here our two hypotheses for each factor or interaction are the presence or absence of an effect of that factor or interaction. For example, a Bayes factor of 10 in favor of an effect would indicate to a logical observer with agnostic prior beliefs that the probability of an effect being present is 10 times greater than the probability there is no effect present. We compute Bayes factors for ANOVAs following the method of Rouder, Morey, Speckman, and Province (2012), using the "BayesFactor" package (Version 0.9.12–2; Morey & Rouder, 2012) for the R statistical computing language. When performing computations, we set the effect size standard deviation to $(\sqrt{2})/2$ and use the defaults within the package for all other settings. We use this standard deviation because it reflects our a priori beliefs about likely effects sizes we may observe if an effect was present in our experiments. Changing this standard deviation within a reasonable range does not greatly affect the conclusions reached (Rouder, Morey, Verhagen, Province, & Wagenmakers, 2016).

To understand how consolidation changes the memory representation maintained in working memory, we applied a process model very similar to that of the Hardman et al. (2017) between-item variant to analyze participant response angle. A diagram of this model is presented in Figure 2. In this modeling framework it is assumed that with some probability, P^M, participants will have an item in memory in order to make their response. With some probability, P^O, this memory will be a fine-detailed memory of the specific angle presented. In this case, responses will be centered on the presented angle with some imprecision, σ^O. With some probability, $(1 - P^O)$, the memory will contain only categorical infor-

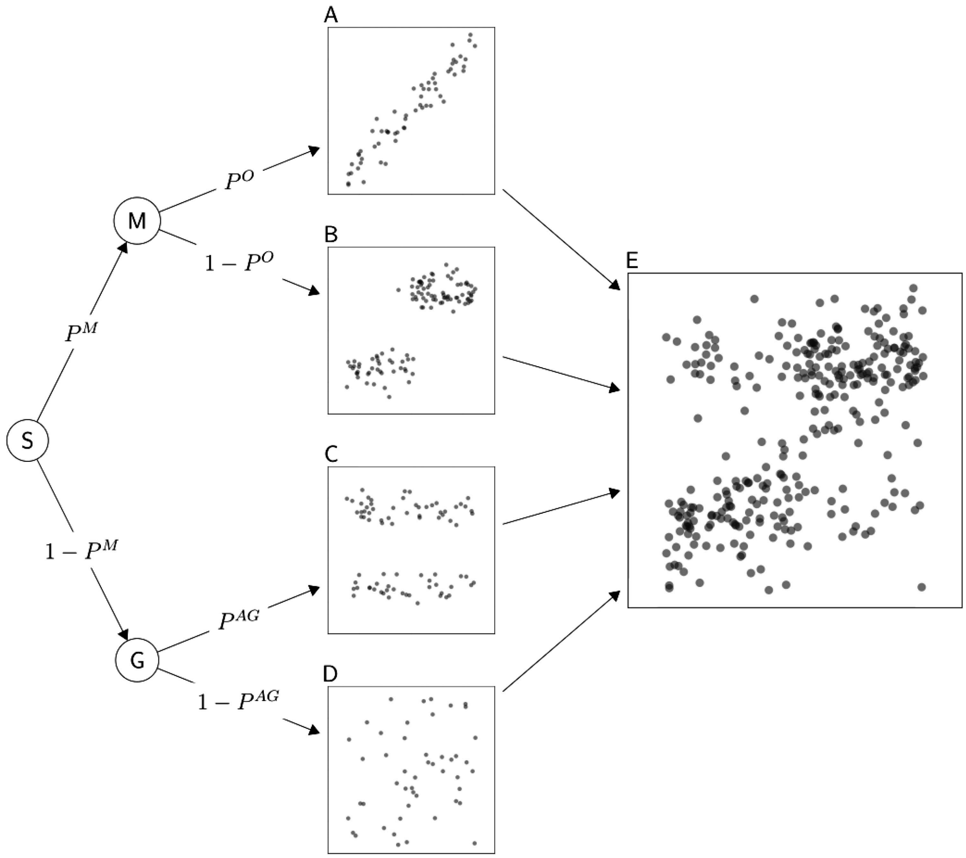

Figure 2. Tree model used to determine the probability of each response distribution in the model of Hardman et al. (2017). P^M = the probability an item is in memory; P^O = the probability that a maintained memory is of the specific angle presented and not categorical in nature; P^{AG} = the probability a guess will be categorical in nature and not random. Each of the panels shows a scatterplot for some example data generated by the model. Angle of stimulus presentation is on the *x* axis and response angle given by the participant is on the *y*-axis in all panels. The continuous memory distribution is shown in Panel A, where fine-grained memory tracks the studied orientation with some imprecision. The categorical memory distribution is shown in Panel B, where the two clusters of responses correspond to remembering that an orientation was to the left or to the right of center, but no more detail. If a participant guesses that the orientation was to the left or the right, it will result in the categorical guessing distribution in Panel C. Finally, uniformly distributed guesses are shown in Panel D. These four individual distributions are combined together into the observed data, shown in Panel E.

mation such as "right." In this case, responses will be centered on some value representative of the category with some amount of noise. The category centers were estimated for each participant. We allowed participants to have up to seven categories, but our analysis showed that participants in our experiment had 3.5 categories on average, typically corresponding to left, right, and center categories. With some probability, $(1 - P^M)$, participants will not have any information about the stimulus in memory. In this case participants must guess. With some probability, P^{AG}, this guess will be a response centered on a random category center with some amount of noise. With probability $(1 - P^{AG})$ this guess will be a random response with a uniform distribution across the response space.

The probability that a response will be drawn from any given distribution follows the tree model described above, detailed in Figure 2. Each of our four types of response states, (1) specific-angle memory, (2) categorical memory, (3) categorical guess, or (4) random guess, are characterized as a distribution or a mixture of several simpler distributions. Each response is assumed to be a sample from a single distribution. Formal model specification, including the specific parameterization of each distribution, can be found in Hardman et al. (2017), although note the modifications discussed in the next paragraph. All distributional probabilities and parameter values are determined through Bayesian Markov chain Monte Carlo (MCMC) sampling techniques. In preliminary analyses we estimated several models without one or more of our four response-state distributions, but fits were extremely poor when any distribution was removed so we do not consider these models further.

We used a model that is conceptually equivalent to the between-item model variant used by Hardman et al. (2017), except for modifications that were required to account for the fact that our stimuli in Experiments 1 through 3 do not exist in a completely circular space, instead only existing in a bounded interval. The main modification was to use truncated normal distributions instead of von Mises distributions at the data level in the model, where the truncation points were at the ends of the range of possible response values. For the weights function and the prior on the category locations, we used nontruncated normal distributions. We did not allow category locations to be outside of the range of response values. To summarize, the von Mises distributions in Equations 3, 4, and 7 of Hardman et al. (2017) were changed to truncated normal distributions and the von Mises distributions in Equations 5 and 20 were changed to normal distributions.

To determine how participant responses change as a function of Consolidation Time and, of secondary interest, Serial Position, we estimated several differing models using this basic framework. Change in participant performance across conditions is captured by estimating effects of experimental condition for specific parameters. Each cell of the experimental design created by crossing Consolidation Time and Serial Position was its own experimental condition for the purpose of model estimation. We have 3 parameters of interest which are good candidates for reflecting the consolidation process, (1) the probability, P^M, that an item was in memory, (2) the probability, P^O, that items in memory were maintained with specific-angle information rather than categorically, and (3) the imprecision of memory, σ^O, when it was maintained as a specific-angle representation. All other parameters were held constant across all conditions.

The model framework of Hardman et al. (2017) is hierarchical in nature, viewing individual participants as samples of the population. As such the model contains a mixture of participant parameters sampled from distributions governed by population parameters. Because our various models differ in overall number of parameters and because the constraint offered by the hierarchical nature of the model framework differs to some extent across models, we compare models using the Watanabe-Akaike Information Criterion (WAIC). This model selection statistic is based on the overall likelihood of the model with a modest penalty term added to adjust for an estimate of the effective number of free parameters (Gelman, Hwang, & Vehtari, 2014). Lower WAIC values indicate better model fit. WAIC is conceptually similar to common fit statistics, such as AIC, but is more appropriate for the model we used. In the context of hierarchical models AIC is problematic because it penalizes models based on the number of free parameters to account for the added flexibility typically afforded by models with more parameters. For hierarchical models like the one we used, the hierarchical parameters are free parameters that actually reduce model flexibility by constraining participant parameters, but are counted against the model in the AIC penalty term. In addition, WAIC is more Bayesian in that it takes the entirety of the posterior variability into consideration. Some fit statistics, such as the deviance information criterion (DIC), instead use the posterior mean of parameters. The model we use contains some parameters that indicate whether or not a category is active and which take on the values 0 or 1. The posterior mean of these parameters will typically be between 0 and 1, but that is not a valid value for these parameters. This quality makes using DIC questionable, but WAIC remains appropriate.

Model-based analysis proceeded in two steps: (1) Model comparison and (2) Hypothesis testing and inference. During model comparison, we determined which factors of the design were relevant for which of the three primary model parameters, excluding any totally irrelevant factors. For our experiments, we had two factors (Consolidation Time and Serial Position), which were crossed with the three parameters, resulting in six factor by parameter terms (e.g., one term would be that P^M is allowed to vary with Consolidation Time).

We used a stepwise procedure in which we started with a full model with all six possible terms in the model, fit that model, and calculated WAIC for that model. Then, we fit all six reduced models that each left out one of the six terms in the full model and calculated WAIC for each of those models. We then selected the model with the best WAIC from these seven models as the best model from this step. If the selected model was the full model, then it was determined to be the best model overall and model comparison was completed. If, instead, one of the reduced models was selected as the best model from that step, the process was repeated with the current step's best model being used as the full model for the next step and the procedure was repeated. For example, if a reduced model with five terms was selected as the best model from the first step then five new reduced models would be fit, each leaving out one term from the new full model. These models then have their WAICs calculated and the model with the best WAIC would be selected. This stepwise procedure continues until the best model selected at a step is the full model for that step.

One way to think about this stepwise procedure is that it successively discards terms. Once a term is discarded it is never considered again. This procedure does not involve fitting all possible models, which means that it is possible for it to miss the best model overall if that best model happens to not be in the search path that the stepwise procedure takes. Its advantage is that it does not require all possible models to be fit while still selecting the best possible model overall with reasonably high probability.

Following the selection of the best model we perform hypothesis tests of main effects and interactions with the remaining factors. This serves to verify which factors and interactions between factors do or do not affect the model parameters. These hypothesis tests are conceptually equivalent to ANOVA and are discussed further in the Appendix. In some cases model selection keeps a condition effect on a parameter but hypothesis testing indicates no evidence for differences in the estimated parameter values across conditions. Most often this will occur when there is tradeoff between two or more parameters, leading to several plausible values and the related ridges in the posterior likelihood. This means that variation is present in the data but it is unclear which parameter can best account for a portion of the variation.

In the context of the present work we use hypothesis tests to determine whether condition effects on individual parameters present in the best-fitting model are clearly leading to the consolidation effect observed in the data. When hypothesis tests cannot clearly confirm differences between parameter values estimated by the best-fitting model then we do not have evidence that these parameters lead to our observed effect. When hypothesis tests do clearly confirm differences between parameter values estimated by the best-fitting model then we have clear evidence that these parameters lead to our observed effect.

All parameter estimation and fit statistics calculation were performed using the "CatContModel" package (Version 0.7.0; Hardman, 2016) for the R statistical computing language. During parameter estimation for the model comparison procedure, we ran 6,000 MCMC chain iterations. The first 1,000 were discarded as burn-in, leaving 5,000 iterations of parameter estimation for use in comparing the models. Once the best model had been selected during model comparison, we ran a further 5,000 iterations of parameter estimation for that model to be used in further analyses (hypothesis tests and parameter values), for a total of 10,000 post-burn-in iterations.

Results

Analysis of mean error. Mean response error, in degrees of angle, is shown in Figure 3 as a function of total Consolidation Time and Serial Position. It is clear from the figure that there was an effect of Consolidation Time. Interestingly the effect is not present for the first serial position. This is statistically confirmed by a Repeated-Measures ANOVA of mean error with Consolidation Time and Serial Position as factors. There was a main effect of Consolidation Time, $F(1, 33) = 88.38$, $\eta_p^2 = .73$, Bayes Factor $= 3.2 * 10^{14}$ in favor of an effect (means: 400 ms = 37.0, 2200 ms = 26.7), a main effect of Serial Position, $F(3, 99) = 83.34$, $\eta_p^2 = .72$, Bayes Factor $= 1.6 * 10^{15}$ in favor of an effect (means: SP1 = 23.2, SP2 = 31.0, SP3 = 35.6, SP4 = 37.4), and an interaction between the two factors, $F(3, 99) = 56.36$, $\eta_p^2 = .63$, Bayes Factor $= 1.1 * 10^{15}$ in favor of an effect.

Model comparison. Having confirmed that a consolidation effect is present, we now explore what process leads to this effect using the model-based approach of Hardman et al. (2017). If consolidation affects the precision of specific-angle memory at test, then we should see that the best fitting model contains differing σ^O parameters for each condition. If consolidation changes the probability that maintained memories contain specific-angle representations relative to coarse categorical representations, then we should see that the best fitting model contains differing P^O parameters for each condition. If consolidation improves the probability that a memory will be maintained at test, then we should see that the best fitting model contains differing P^M parameters for each condition.

Table 1 lists the model fit, as indexed by WAIC, for all of the models we tested using our stepwise procedure. The full model

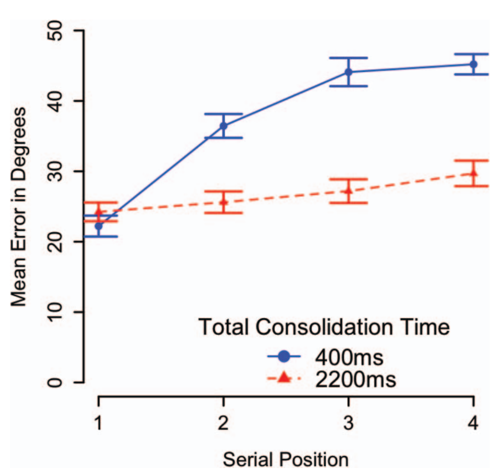

Figure 3. Mean response error in Experiment 1 as a function of Consolidation Time and Serial Position. Error bars represent standard error of the mean. See the online article for the color version of this figure.

Table 1
WAIC for All Tested Models in Experiment 1

Model	WAIC	Difference from best model
Full model	144,313	0
Constant P^M across serial position	144,645	332
Constant P^M across consolidation time	145,481	1167
Constant σ^O across serial position	144,408	95
Constant σ^O across consolidation time	144,380	66
Constant P^O across serial position	144,402	88
Constant P^O across consolidation time	144,467	153

Note. WAIC = Watanabe-Akaike Information Criterion. In the full model parameters were estimated freely for all Serial Positions and Consolidation Times. In our stepwise procedure we estimated the full model but constrained a single parameter value to be invariant across Serial Position or Consolidation Time. WAIC and difference scores are rounded to the nearest whole number. p^M = the probability an item is in memory; σ^O = the standard deviation of the specific-angle response; p^O = the probability that a maintained memory is of the specific angle presented and not categorical in nature.

containing effects of consolidation time and serial position on all target parameters was the clear winner beating the next best-fitting model by 63 points. Table 2 lists all estimates of the target parameters under the full model as a function of consolidation time and serial position.

Figure 4 shows both participant responses and simulated data using the parameter estimates of our best-fitting model. Inspection of Figure 4 clearly indicates that the model does a good job of characterizing the major patterns in the data. All of the major patterns in data are replicated in the model. There is one small difference between the participant and model-simulated data sets. Participants respond with smaller variability around the central category center than around peripheral left and right category centers while the model uses a common variability parameter for all categories. We constrain category variability to maintain model identifiability, but this means the model does necessarily miss such a difference. Despite this small misspecification, it is clear from Figure 4 that our model has good fit, producing data representative of participant performance.

Bayes factor analysis of parameter values. The best model selected during model comparison contained all possible terms (factor by parameter combinations), so we now perform hypothesis tests for main effects of Consolidation Time and Serial Position and their interaction for all three primary parameters. These tests are conceptually equivalent to the like-named ANOVA tests and give us extra information beyond what can be gleaned from model comparison. See the Appendix for more information about how these tests were performed. To summarize the results, we found that all main effects and interactions were present, except for the interaction between Consolidation Time and Serial Position for the P^O parameter. The parameter means are given in Table 2 and complete hypothesis test results are given below.

For the probability that an item was in WM (P^M), we found a main effect of Consolidation Time, Bayes factor $= 5.7 * 10^{18}$ in favor of an effect, a main effect of Serial Position, Bayes factor $= 2.7 * 10^{12}$ in favor of an effect, and an interaction between Consolidation Time and Serial Position, Bayes factor $= 7.8 * 10^4$ in favor of an effect. The pattern is essentially that P^M drops from

Table 2
Mean Parameter Estimates for the Best-Fitting Model in Each Experiment

	P^M				σ^O				P^O			
Consolidation condition	SP1	SP2	SP3	SP4	SP1	SP2	SP3	SP4	SP1	SP2	SP3	SP4
Exp 1												
Shortest	.85	.46	.31	.30	15.1	19.6	25.1	34.3	.33	.35	.54	.61
Longest	.81	.78	.73	.71	17.0	18.8	17.6	21.1	.26	.32	.33	.45
Exp 2												
Shortest	.86	.28	.14	.08	16.4	16.4	16.4	16.4	.33	.22	.41	.66
Longest	.72	.61	.50	.51	16.4	16.4	16.4	16.4	.26	.43	.40	.52
Exp 4												
Shortest	N/A	.11	.06	.02	N/A	25.0	23.9	30.5	N/A	.28	.43	.57
Longest	N/A	.42	.29	.16	N/A	22.2	27.5	38.4	N/A	.68	.42	.76

Note. SP = Serial Positio; P^M = the probability an item is in memory; σ^O = the standard deviation of the specific-angle response; P^O = the probability that a maintained memory is of the specific angle presented and not categorical in nature. In Experiment 2 the model was fit by binning trials into 200 ms consolidation duration bins. The Experiment 2 "Shortest" condition in this table refers to the bin containing Consolidation Times of 200–400 ms. The Experiment 2 "Longest" condition in this table refers to the bin containing Consolidation Times of 1,000–1,200 ms. For Experiment 4, SP1 values are not applicable because we only modeled items that followed a manipulated consolidation time.

Serial Position 1 to Serial Position 4, but is hurt substantially more in the short Consolidation Time condition than the long Consolidation Time condition.

For the imprecision of continuous items in WM (σ^O), we found a main effect of Consolidation Time, Bayes factor = 22 in favor of an effect, a main effect of Serial Position, Bayes factor = 57 in favor of an effect, and an interaction between Consolidation Time and Serial Position, Bayes factor = 10 in favor of an effect. The pattern of means is that σ^O increases with Serial Position, but increases substantially more in the short Consolidation Time condition than the long Consolidation Time condition.

Finally, for the probability that an item was stored continuously in WM (P^O), we found a main effect of Consolidation Time, Bayes factor = 14 in favor of an effect and a main effect of Serial Position, Bayes factor = 8.8 in favor of an effect. There was no interaction, however, with Bayes factor = 110 against an effect. The main effects are that P^O increases with Serial Position and decreases with Consolidation Time.

Discussion

The results of our first experiment replicate previous research, showing that more time for consolidation can improve working memory performance even after masking (Nieuwenstein & Wyble, 2014; Ricker & Cowan, 2014). One unexpected finding was that no consolidation effect was present at serial position 1. Instead, performance was similar at serial position 1 regardless of available time for consolidation. This similarity coupled with the relatively good performance indicates that consolidation of the first item was always completed regardless of the time given to do so. This conclusion is supported by the lack of a traditional bow-shaped serial position curve with the long consolidation time. Instead, response error was very similar in size across all serial positions. This may indicate that all items are fully consolidated into memory under our long consolidation condition. We explore this further in Experiments 2 and 3.

The similarity in response accuracy for serial position 1 in both consolidation conditions deserves some further comment. There are several possible explanations of this finding, but we favor the following process as a likely description. Whenever the consolidation process was engaged it was always completed, at least in the context of this experiment. The deficit we see for serial positions 2–4 in the short consolidation condition would then be attributable to a delay in switching attention to the new item when it appears on screen. When still engaged in consolidation of the previous item when a new item appears, switching attention to the new item is delayed just as switching attention to a secondary task is delayed when consolidation is ongoing (Jolicœur & Dell'Acqua, 1998; Nieuwenstein & Wyble, 2014; Ouimet & Jolicoeur, 2007; Stevanovski & Jolicœur, 2007). This leaves less time for encoding and consolidation of the new item because of the delayed shift, resulting in larger errors.

This explanation of our findings is consistent with an attentional blink explanation of the consolidation effect. If an attentional blink is occurring in our procedure, then processing of items after serial position 1 suffers when consolidation is not completed for the previous item at the time of onset of the new memory item. This finding of no consolidation effect for the first serial position is also problematic for explanations of the consolidation effect that describe the benefit of more consolidation time as an improvement in memory performance for the preceding memory item. More time for consolidation did not improve performance for the first item.

The results of our model-based and Bayes factor analyses are also consistent with the interpretation that other factors also play a role in addition to an attentional blink. Stepwise model comparison indicated that all three of our target parameters change as a function of consolidation time. Likewise Bayesian hypothesis testing indicates that there were differences in the estimated parameter values across conditions for all three target parameters.

Although our tests indicate that all parameters play a role in the consolidation effect, the biggest driver of the effect was the P^M, the probability that an item was in memory. Our estimates of P^M changed by 0.28, out of a total range of 1.0, across conditions. Changes in our other two parameters of interest were much more modest. The imprecision of the specific-angle response, σ^O, only changed by 5 degrees across conditions. The probability of maintaining a continuous representation given that a memory was present, P^O, only changed by 0.12 across conditions. It's important to note that a change

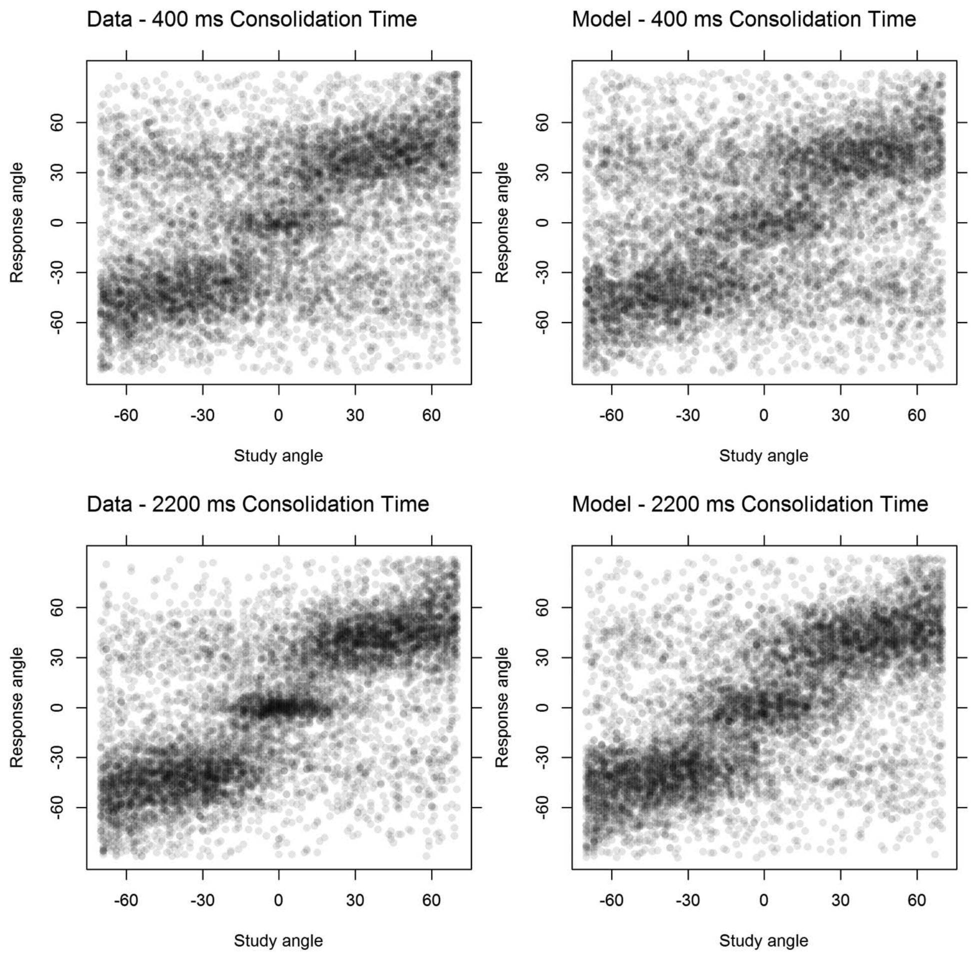

Figure 4. Response angles produced in Experiment 1 as a function of the stimulus angle presented at study for all participants. The top-left panel shows experimental data from the 400 ms condition while the bottom-left panel shows experimental data from the 2,200 ms condition. The top-right panel shows simulated data using the parameters of the best-fitting model from the 400 ms condition while the bottom-right panel shows simulated data using the parameters of the best-fitting model from the 2,200 ms condition.

in P^M has a much larger impact on errors than does an equivalent change in P^O due to the nature of the parameters.

The results of Experiment 1 indicate that at least 3 processes are ongoing during consolidation: increases in representational precision, qualitative change in the nature of the maintained representation, and a change in the probability that memory items are present in memory at test. The first two processes appear to have a relatively minor contribution to changes in performance and the mean performance across serial positions seems to confirm this interpretation. Improvement in performance did not occur for the first item presented, seeming to indicate that a proceeding item had to be present for the consolidation time to affect performance. This is more consistent with either an attentional blink or temporal distinctiveness process than with a gradual improvement in memory.

Experiment 2

Having found that the consolidation effect is largely driven by a change in the probability an item is in memory at test, we now turn to understanding what process is leading to this change. In Experiments 3 we investigate whether we are observing temporal distinctiveness effects, but first we must understand the time course of the consolidation process at work in Experiment 1. This time course is crucial to uncovering whether temporal distinctiveness is the underlying cause of the improvement in working memory performance with more time for consolidation. We explore the question of how long short-term consolidation lasts in our present paradigm in Experiment 2.

Past work on the time course of consolidation seems to limit it to under a few seconds in duration, but reports often conflict on more precise estimates (Jolicœur & Dell'Acqua, 1998; Lagroix et al., 2012; Nieuwenstein, Potter, et al., 2009; Nieuwenstein, Van der Burg, Theeuwes, Wyble, & Potter, 2009; Nieuwenstein & Wyble, 2014; Stevanovski & Jolicœur, 2007). Much of the working memory literature on consolidation has used secondary task reaction time (RT) to infer the time course (Jolicœur & Dell'Acqua, 1998; Stevanovski & Jolicœur, 2007), further obscuring the actual time it takes to consolidate a working memory trace. While RT is a valuable tool for understanding psychological process, using RT measures from a second task to infer specific memory processes in the first is indirect and could be confounded with other processes. It could also be that consolidation durations vary across stimulus categories (Jolicœur & Dell'Acqua, 1998; Nieuwenstein & Wyble, 2014). For example, more complex memory items may take longer to consolidate than simple ones. For these reasons understanding the time course of consolidation in the present paradigm requires us to measure it directly.

In the present experiment we use the approach from Experiment 1 but vary the postmasking period between item presentations by increments of ~17 ms up to a maximum time of 1000 ms to understand the precise time course of the consolidation function. If we reach a time point after which further time for consolidation does not improve working memory performance we will have determined the duration of the consolidation process for the present paradigm and stimuli.

Method

Participants. Twenty-five students (aged 18 – 26, 12 female) enrolled in introductory psychology at the College of Staten Island participated in exchange for partial course credit. All participants had normal or corrected-to-normal vision. This sample size was arrived at using the same protocol as in Experiment 1.

Materials. All materials were the same as in Experiment 1, but with a small difference in size attributable to the use of different computer monitors. In this experiment the ring was 2.3 cm in diameter and the dot was 0.3 cm in diameter.

Design. The design differed from Experiment 1 in that the consolidation period varied randomly in duration between 17 and 1000 ms. Total consolidation period length could take any value within this range, including the end points, with the constraint that the duration had to be a multiple of the screen refresh rate of 17 ms. Participants completed 8 practice trials followed by 6 blocks of 30 experimental trials. Each trial contained 4 delayed estimation responses resulting in 720 total experimental observations per participant.

Procedure. The procedure was the same as in Experiment 1 except for the change in consolidation period duration. Consolidation period duration was still manipulated between trials, with all consolidation periods on a single trial being equal in length. Because different computer monitors were used in this experiment, the memory stimulus presentation locations were now 4.2 cm from the center of the screen.

Analysis. In this experiment we use a linear models approach that is a mixture of multiple regression and ANOVA. We then calculate Bayes factors for individual variables in the linear model with the "BayesFactor" package (Version 0.9.12–2; Morey & Rouder, 2012) for the R statistical computing language. When performing computations, we set the effect size standard deviation to $(\sqrt{2})/2$ for fixed factor effects and 1/2 for slopes. We use the defaults within the package for all other settings. To calculate Bayes factors for main effects we compare the model with both main effects of interest (Consolidation Duration and Serial Position) to a model without one of the main effects. To calculate Bayes factors for the interaction between our main effects we compare a model with both main effects and an interaction to a model with only the main effects. If the model including the effect in question results in a higher probability of generating the observed data than does the model without it, we will observe a Bayes factor in favor of the effect. If the opposite is true, then we will observe a Bayes factor in favor of the null model, meaning that the effect of interest is not likely to be present. In this manner we can investigate how adding effects of interest improves our model of the observed data.

Results

Analysis of mean error. Loess regression (Cleveland, Grosse, & Shyu, 1992) of overall performance, measured as error in degrees of angle, is plotted as a function of Consolidation Time in Figure 5 for each serial position individually. It is clear that there was a large effect of Consolidation Time, as expected from Experiment 1 and in agreement with previous research (Nieuwenstein & Wyble, 2014; Ricker & Cowan, 2014). As in Experiment 1, Consolidation Time improved performance for all serial positions except for serial position 1. Instead, for serial position 1, more time for consolidation led to larger errors relative to a shorter consolidation period.

To explore the effect of Consolidation Time and any interaction with Serial Position on response error, we constructed a linear model that contains factors varying in discrete levels (Participant and Serial Position) and a covariate that varied continuously across its range (Consolidation Time). As we are not interested in individual participant differences, we follow the approach used in Repeated-Measures ANOVA and treat each participant as a random main effect that does not interact with any other factors. Following the method described in the analysis section above, we verified that both main effects are statistically supported, Consolidation Time Bayes factor $= 6.1 * 10^{16}$ in favor of an effect, Serial Position Bayes factor $= 4.3 * 10^{77}$ in favor of an effect, as is the interaction of these two effects, Consolidation Time $*$ Serial Position Bayes factor $= 5.0 * 10^{17}$ in favor of an effect.

The nature of the interaction can be understood by examining the linear regression beta weights for the consolidation function at

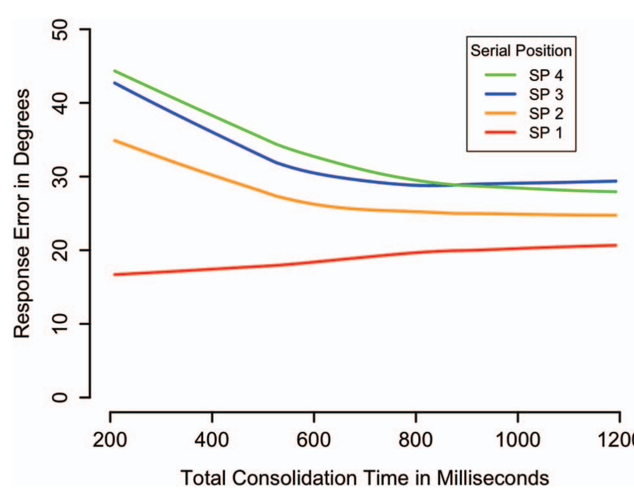

Figure 5. SP = Serial Positio. Results of Experiment 2 presented as a loess regression of Consolidation Time on response error for each serial position individually. See the online article for the color version of this figure.

each individual serial position (Serial Position beta weights [95% confidence interval]; 1 = +7.69 [+4.92, +10.46], 2 = −11.45 [−8.02, −14.87], 3 = −13.03 [−9.40, −16.67], 4 = −15.56 [−11.84, −19.28]). Although the beta weight is positive for Serial Position 1, meaning larger errors with more time for consolidation, the beta weights are negative for all other serial positions, indicating smaller errors with more consolidation time. This replicates the finding from Experiment 1 that the first serial position does not benefit from increased time available for consolidation.

Although the loess regression in Figure 5 provides a good visual representation of the data, it is difficult to see where the function asymptotes. Figure 6 uses the following method to create a clear visual representation of the consolidation function asymptote. To better determine the time needed to fully consolidate an item, we binned each trial into one of 10 bins based upon the length of the postmasking consolidation period. Each bin contained trial data across a 100 ms period: 0-100 ms, 101-200 ms, and so forth. This was done only for Serial Positions 2 – 4 as there was no benefit to consolidation at Serial Position 1. We then examined mean error across these bins to see when performance reached asymptotic levels. As can be seen in Figure 6, mean error decreased for roughly the first 600 ms (150 ms presentation + 50 ms mask + 400 ms postmasking consolidation period) after stimulus onset. After 600 ms mean error no longer decreased.

Model comparison. Having confirmed that a consolidation effect is present, we now explore what process leads to this effect. The Hardman et al. (2017) model only supports categorical conditions (or factor levels), not continuous covariates, which means that the consolidation times had to be categorized for model fitting. Thus, we aggregated consolidation times into 5 bins that were each 200 ms wide and treated these bins as conditions for the purpose of parameter estimation. These bin widths seemed sufficient to capture the patterns in the data, which were fairly smooth and gradual with respect to consolidation time. Table 3 lists model fit,

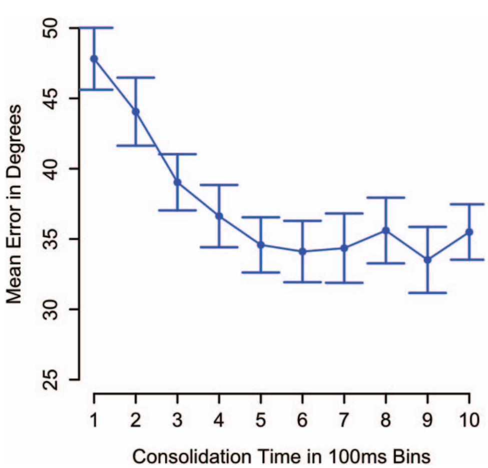

Figure 6. Mean response error (in degrees) in Experiment 2 as a function of total Consolidation Time binned in 100 ms increments. For example, bin 2 contains all data with Consolidation Times between 301–400 ms (Presentation Duration + Mask Duration + Post-Masking Consolidation Period Duration). Data from serial position 1 is excluded because it did not demonstrate a consolidation effect. Error bars represent standard error of the mean. See the online article for the color version of this figure.

Table 3
WAIC for All Tested Models in Experiment 2

Model	WAIC	Difference from best model
Full model	164,310	53
Constant P^M across serial position	165,283	1026
Constant P^M across consolidation time	165,007	750
Constant σ^O across serial position	164,278	21
Constant σ^O across consolidation time	164,436	179
Constant P^O across serial position	164,668	411
Constant P^O across consolidation time	164,449	192
Constant σ^O across serial position and . . .		
Constant P^M across serial position	164,691	434
Constant P^M across consolidation time	164,716	459
Constant σ^O across consolidation time	164,257	0
Constant P^O across serial position	165,303	1046
Constant P^O across consolidation time	165,314	1057
Constant σ^O across serial position & consolidation time and . . .		
Constant P^M across serial position	164,546	289
Constant P^M across consolidation time	164,613	356
Constant P^O across serial position	165,321	1064
Constant P^O across consolidation time	165,308	1051

Note. WAIC = Watanabe-Akaike Information Criterion. In the full model parameters were estimated freely for all Serial Positions and Consolidation Times. In our stepwise procedure we estimated the full model but constrained a single parameter value to be invariant across Serial Position or Consolidation Time. In the case that the full model was not the best-fitting model, we kept the best model and repeated the procedure. WAIC and difference scores are rounded to the nearest whole number. P^M = the probability an item is in memory; σ^O = the standard deviation of the specific-angle response; P^O = the probability that a maintained memory is of the specific angle presented and not categorical in nature.

as indexed by WAIC, for all of the models we tested in our stepwise procedure. The model containing effects of consolidation time and serial position on P^M and P^O, the probabilities that an item was in memory and maintained as a specific-angle memory respectively, but no effect of consolidation time or serial position on σ^O, the imprecision of the specific-angle response, was the winner beating the full model by 53 points. Table 2 lists all estimates of the target parameters under the full and best-fitting models as a function of consolidation time and serial position.

Bayes factor analysis of parameter values. The best model selected during model comparison only included effects of Consolidation Time (binned into 5 bins) and Serial Position for the P^M and P^O parameters. Following this model we only test main effects and interactions for those two parameters and not for σ^O. The parameter means for the shortest and longest Consolidation Times can be found in Table 2.

For the P^M parameter, we found a main effect of Consolidation Time, Bayes factor = $1.5 * 10^8$ in favor of an effect, a main effect of Serial Position, Bayes factor = $4.1 * 10^{10}$ in favor of an effect, and an interaction between Consolidation Time and Serial Position, Bayes factor = 80 in favor of an effect. As Serial Position increases, P^M decreases at some rate, but the nature of the interaction is that the rate of decrease is faster when Consolidation Time is shorter.

For the P^O parameter, we found a main effect of Serial Position, Bayes factor = 6.3. We found no main effect of Consolidation Time, Bayes factor = 20 against an effect, and no interaction, Bayes factor = 10 against an effect. The main effect of Serial Position is that P^O increases as Serial Position increases.

Discussion

The results of Experiment 2 show that consolidation within visual working memory is a fast process completed in well under one second. Here we directly observe when consolidation of memory items was completed instead of inferring it through another task, making interpretation of the present results straightforward. Our estimate is that the entire process is complete within 600 ms is in agreement with the estimates of the length of the attentional blink (Lagroix et al., 2012; Nieuwenstein, Potter, et al., 2009). This provides more evidence that giving more time for consolidation of a working memory trace may improve performance by helping to avoid the attentional blink.

Our modeling analysis again provided evidence that the main driver of the consolidation effect is the probability that an item is in memory at test. In contrast with Experiment 1, here we found evidence against a qualitative change in the nature of the memory maintained, P^O, or a change in memory imprecision, σ^O. As in Experiment 1, the change in the probability that an item will be in memory could be due to either an attentional blink process or temporal distinctiveness effects.

Experiment 3

In Experiment 3 we investigate whether working memory consolidation or temporal distinctiveness is responsible for the pattern of results in Experiments 1 and 2. Temporal distinctiveness approaches to memory posit different memory mechanisms than do working memory and short-term consolidation approaches. If temporal distinctiveness can account for the experimental effect in the present paradigm it would indicate that there is no need to hypothesize a short-term consolidation process.

Temporal Distinctiveness Theory (Brown et al., 2007; Crowder, 1976; Souza & Oberauer, 2015; Shipstead & Engle, 2013) posits that all memory performance can be understood in terms of discriminability. According to this approach, time is an important factor in determining discriminability because it is used as a retrieval cue when making a response. Items encoded farther apart in time from one another have more distinct retrieval cues and are therefore more discriminable from one another at recall, leading to better overall performance. In the present context, temporal distinctiveness approaches can explain our results by arguing that shorter consolidation periods impair performance not because of any consolidation process but because they lead to less distinct memories.

A central element to temporal distinctiveness approaches to memory performance is scale independence. This means that you should observe discriminability-based memory impairment across all time periods, whether 1 second or 1 week, so long as the temporal discriminability ratios stay constant. Following a straight ratio rule, there should be similar levels of temporal discriminability with a target that occurred 7 weeks ago coupled with a distractor that occurred 5 weeks ago, as there is with a target that occurred 7 seconds ago coupled with a distractor 5 seconds ago. In the context of our present work, if distinctiveness is driving the time-based improvement in memory then performance will improve whenever the temporal distinctiveness ratios are increased, regardless of scale. We manipulate these distinctiveness ratios just as we would the time available for consolidation, by manipulating the amount of time between item presentations.

Our estimates of the time course of short-term consolidation are not based upon scale independence and bring us to differing predictions from temporal distinctness theory in many circumstances. Following our short-term consolidation hypothesis, items require a certain period of time after presentation for consolidation in visual working memory. In the case of the stimuli used here this time should be roughly 600 ms, given the results of Experiment 2. Once this period of time has passed, consolidation is complete and further time provided is not beneficial. If a consolidation process is driving our temporal effect, then performance will only improve over the initial 600 ms after stimulus onset. Additional time separating the studied items beyond 600 ms is not predicted to have an effect following the consolidation approach.

In Experiment 3 we continue to use the same experimental paradigm as in Experiment 2, but we lengthen all events during the memory item presentation sequence such that the time for consolidation in Experiment 3 ranges from 767–2767 ms depending on the length of the postmask consolidation period. Because our shortest consolidation period in this experiment is longer than the time needed to complete the consolidation process that we hypothesize, our consolidation approach predicts no effect of the additional time provided between items in Experiment 3. Unlike a consolidation process, a temporal distinctiveness approach has no theoretical reason to predict a hard cap on the benefit of additional time between item presentations. In order to be sure that temporal distinctiveness approaches predict an effect of additional time between items in Experiment 3, we must first calculate the temporal ratios discussed in the previous paragraphs and use them to determine each item's distinctiveness. Higher distinctiveness for a given item is equivalent to predicting better performance on correct recall of that item. If calculated distinctiveness increases with longer SOAs between memory items in Experiment 3, then we would expect to see increased accuracies with longer SOAs in this experiment.

How temporal ratios are computed and used to determine distinctiveness differs slightly from author to author, but a generally accepted mathematical model of computing distinctiveness derived from these ratio rules is detailed by Brown et al. (2007). We follow the approach of Brown et al. in determining temporal distinctiveness in Experiment 3. An important element of this approach is that time is assumed to be logarithmically compressed such that temporal distances in the distant past seem closer together than the same distances in the recent past. This is functionally implemented by raising the temporal ratios to the power of a scaling parameter, c.

Distinctiveness measures for each serial position, computed following the method of Brown et al. (2007), are given in Table 4 for the shortest and longest consolidation period in this range.[1] We provide distinctiveness measures under a small compression value,

[1] When calculating temporal distances we considered the time of memory probe onset to be the time of memory recall. To account for the time required to make prior responses we added the mean response time to all temporal distances for each prior serial position recalled. For example, when calculating the temporal distance between presentation of the first item and the third item probe we calculated, (time between offset of serial position 1 presentation to onset of serial position 1 probe) + (2 $*$ mean response time).

Table 4
Temporal Distinctiveness for Each Serial Position as a Function of Consolidation Time and Scaling Parameter C

Consolidation condition	SP1 (c = 2)	SP2 (c = 2)	SP3 (c = 2)	SP4 (c = 2)	SP1 (c = 12)	SP2 (c = 12)	SP3 (c = 12)	SP4 (c = 12)
Experiment 2								
Shortest	1.5	.49	.42	.47	74.98	2.48	1.22	2.06
Longest	1.2	.72	.64	.93	35.73	12.19	12.19	43.51
Experiment 3								
Shortest	1.41	.67	.55	.74	58.29	9.11	5.69	15.14
Longest	1.21	.81	.79	1.31	36.51	18.02	32.58	212.31

Note. SP = Serial Positio. Scaling parameter c indexes the compression of latent temporal space. Values of c here are typical of those used to fit data in Brown et al. (2007). Following temporal distinctiveness approaches, larger distinctiveness values should result in better memory performance (lower mean error).

c = 2, and a large one, c = 12, both of which are typical of the values used by Brown et al. In both cases there is a clear pattern of changing distinctiveness values from the short to long consolidation time. This means that in Experiment 3 there should be an effect of Consolidation Time if temporal distinctiveness is leading to the effect observed in Experiments 1 and 2. To reiterate, if consolidation of visual working memory traces is producing our effect, we should see no improvement in error rates with longer Consolidation Time because consolidation of all items should be complete before the end of our shortest consolidation period.

Method

Participants. Twenty-six students (aged 18 – 29, 15 female) enrolled in introductory psychology at the University of Missouri participated in exchange for partial course credit. All participants had normal or corrected-to-normal vision. This sample size was arrived at using the same protocol as in Experiment 1.

Materials. All materials were the same as in Experiment 1.

Design. The design differed from Experiment 2 in that the range of postmasking consolidation period durations varied randomly between 167 and 2167 ms, resulting in total Consolidation Times of 767–2767 ms. Participants completed 8 practice trials followed by 4 blocks of 30 experimental trials. Each trial contained 4 delayed estimation responses resulting in 480 total experimental observations per participant. In all other ways the design was the same as Experiment 2.

Procedure. The procedure was the same as in Experiment 2 with the following exceptions. In Experiment 3, memory items were displayed for 400 ms, masking stimuli were displayed for 200 ms, and the consolidation period duration varied randomly between 167 and 2167 ms. Stimulus measurements followed those of Experiment 1 as the same computer screens were used for both experiments.

Analysis. The analysis uses the same linear model and method described in Experiment 2.

Results

Loess regression of overall performance, measured as error in degrees of angle, is plotted as a function of Consolidation Time in Figure 7 for each serial position individually. It is clear that there was no effect of Consolidation Time, as expected following the results of Experiment 2. To confirm this lack of effect statistically we again constructed a linear model with the same structure as in Experiment 2 (Factors: Participant, Consolidation Time, and Serial Position). Following the method described in the analysis section of Experiment 2, we verified that there was no effect of Consolidation Time, Bayes factor = 8.5 in favor of the null, an effect of Serial Position, Bayes factor = $6.8 * 10^{27}$ in favor of an effect, and no interaction of these two effects, Bayes factor = 28.1 in favor of the null (Serial Position beta weights [95% confidence interval]; 1 = −0.37 [−1.64, +0.91], 2 = −0.25 [−1.59, +1.10], 3 = −1.55 [−3.26, +0.11], 4 = −0.68 [−2.48, +1.12]).

We do not report the results of our computational modeling analysis in Experiment 3 because there was no effect present. Without an effect there are no data to drive changes in parameter values across conditions. To confirm this assumption we fit our model to the data from this experiment and found that the only effect present was a change in P^M, the probability an item was in memory, as a function of serial position, Bayes factor = $5.8 * 10^4$ (all other Bayes factors greater than 4.9 in favor of no effect being present).

Discussion

The model of temporal distinctiveness from Brown et al. (2007) predicts an effect of Consolidation Time in the present experiment

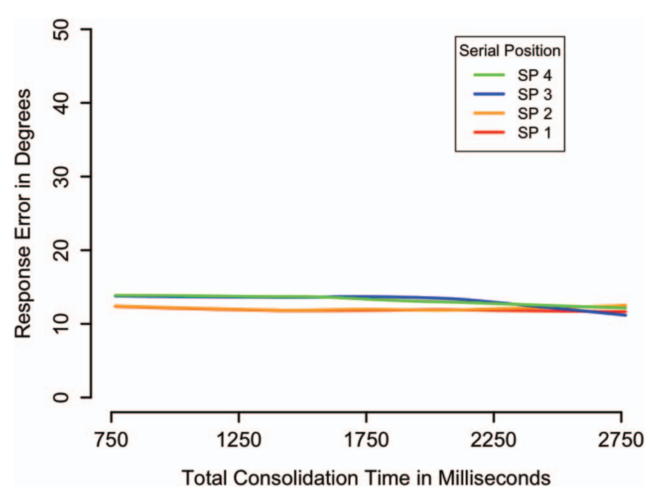

Figure 7. SP = Serial Positio. Results of Experiment 3 presented as a loess regression of Consolidation Time on response error for each serial position individually. See the online article for the color version of this figure.

(see Table 4) yet none was observed. In contrast, our theory of working memory consolidation coupled with the results of Experiment 2 led us to predict that there would be no effect of Consolidation Time in the present experiment. All Consolidation Times used were longer than that needed to fully consolidate the memory items. The consolidation approach predicts that when items are fully consolidated, providing more time before presentation of the next item should have no effect.

Experiment 3 confirms the predictions of the consolidation approach, providing evidence against an effect of varying Consolidation Time between 767–2767 ms. This is in conflict with temporal distinctiveness explanations of our results but in harmony with consolidation explanations. Similarly, Ricker and Cowan (2014) demonstrated an effect of working memory consolidation in visual working memory change detection of unfamiliar figures while Ricker, Spiegel, and Cowan (2014) ruled out an effect of temporal distinctiveness in the same paradigm. Converging evidence from multiple approaches, recall and change detection, using differing memory stimuli, orientation and unfamiliar figures, confirm that there is a consolidation effect within visual working memory, while also providing strong evidence against the presence of temporal distinctiveness effects.

Experiment 4

Having ruled out temporal distinctiveness effects as leading to our consolidation effect, we now turn to testing our alternative hypothesis that an attentional blink is responsible. The attentional blink approach to understanding consolidation effects predicts an increase in the probability an item is in memory at the time of test when more time is given for consolidation. This is what we have observed as the driving cause of increased performance in Experiments 1 and 2. In this experiment we test a stronger prediction of the attentional blink approach to understanding working memory consolidation. We selectively manipulate the consolidation period following a single item on each trial and look for selective signatures of an attentional blink process.

Modern mathematical models of the attentional blink posit that the blink is caused by the inability to switch attention to a new target during the period that comprises the blink (Wyble et al., 2009, 2011). In these models consolidation of a memory item is always completed once it begins. Additionally, consolidation into working memory suppresses attention while the process is ongoing. If these models are correct then we should expect to see that items *immediately following* a consolidation period are strongly affected by the length of that period, with shorter consolidation times leading to lower probabilities that the immediately following item is in memory. This is because attention would be unable to shift to the new memory item when it is presented within the attentional blink period of the previous item. In contrast, items *immediately preceding* a consolidation period should be unaffected by the duration of the consolidation period because consolidation is always completed once it begins.

A model of consolidation similar to this attentional blink approach but still allowing for the reallocation of attention when a new item appears could also account for much of the consolidation effect we have observed. We call this the flexible attention approach. Consolidation of a memory item would be assumed to fully occupy attention and successful completion would require a fixed amount of time. If attention does not dwell on the memory item for the required amount of time then the sensory trace is not stabilized and the item is not consolidated. The flexible attention approach differs from the attentional blink approach in that attention is switched to a new item when it appears in the sequence and consolidation of the new item begins immediately.

Following the flexible attention theory, we should see that items *immediately following* a consolidation period are unaffected by the length of that period. This is because attention would shift to the new memory item when it is presented. In contrast, items *immediately preceding* a consolidation period should be strongly affected by the duration of the consolidation period, with shorter consolidation times leading to lower probabilities that an item is in memory. This is because attention is more likely to shift away from the preceding item before consolidation is complete.

We can also imagine several theories that would predict that a longer consolidation period would affect all items on the list, not only the items immediately preceding or following the manipulated consolidation period. For example, if consolidation functions as a critical period during which attention-based maintenance is especially important (Barrouillet, Plancher, Guida, & Camos, 2013) then we should see that all items currently in memory are sensitive to the length of any single consolidation period.

To summarize, Experiment 4 is a replication of Experiment 1, except that only a single consolidation period is manipulated on each trial. In Experiment 1 all consolidation periods were manipulated on each trial. The attentional blink approach predicts that we should see a selective influence of the manipulated consolidation period on the item immediately following it. For example, if the second consolidation period is manipulated then errors should change for the third item in the list. The flexible attention approach predicts that we should see a selective influence of the manipulated consolidation period on the item immediately preceding it. For example, if the second consolidation period is manipulated then errors should change for the second item in the list.

Method

Participants. Forty-nine students (aged 18 – 34, 33 female) enrolled in introductory psychology at the College of Staten Island participated in exchange for partial course credit. All participants had normal or corrected-to-normal vision. This sample size was arrived at using the same protocol as in Experiment 1, except that the minimum number of participants was 40. The increase in target sample size is attributable to a decrease in the number of trials that contribute to critical condition means.

Materials. All materials were the same as in Experiment 2, except that the dot could appear on the full range of the circle. Because of this change the range of participant errors can be much larger without necessarily implying lower performance as compared to previous experiments.

Design. The design was identical to Experiment 1 except for the following differences. In the present experiment the consolidation period duration following each item was held constant at 200 ms except for the period following one item on each trial. The manipulated duration was selected at random from the following values 17, 200, or 500 ms, resulting in total Consolidation Times of 217, 400, or 700 ms. The serial position of the manipulated consolidation duration was selected at random on each trial, with serial position 1 referring to the consolidation duration following the first memory item. Participants completed 8 practice trials followed by 6 blocks of 30 experimental trials. Each trial contained 4 delayed estimation responses resulting in 720 total experimental observations per participant.

Procedure. The procedure was the same as in Experiment 1 except for the change in the way the consolidation period duration was manipulated. In this experiment only the consolidation period following a single memory item was manipulated on each trial, while all other consolidation periods were held constant at 200 ms.

Analysis. The analysis is the same as that used in Experiment 1, except that the mathematical model used was the version of the model described in Hardman et al. (2017) for use with circular data. In Experiments 1 through 3 the data were in a linear space and the model specification reflected this constraint.

Results

Analysis of mean error. Mean response error, in degrees of angle, is shown in Figure 8 as a function of total Consolidation Time and Serial Position. The figure shows an effect of consolidation time that was isolated to the serial position immediately following the manipulated period. This can be seen more clearly in Figure 9 which shows the effect of consolidation on the items immediately before (panel a) and immediately after (panel b) the consolidation period manipulated. It is clear that there is no effect of consolidation time in panel a and a strong effect of consolidation time in panel b.

This is statistically confirmed by two Repeated-Measures ANOVAs of mean error with Consolidation Time and Serial Position as factors. The first ANOVA was conducted using only those items which immediately preceded the manipulated consolidation position. In this analysis there was no effect of Consolidation Time, $F(2, 96) = 2.18$, $\eta_p^2 = .04$, Bayes Factor = 38 in favor of the null (means: 217 ms = 74.3, 400 ms = 72.0, 700 ms = 74.5), a main effect of Serial Position, $F(3, 144) = 55.28$, $\eta_p^2 = .54$, Bayes Factor $= 2.0 * 10^{46}$ in favor of an effect (means: SP1 = 55.9, SP2 = 72.7, SP3 = 81.0, SP4 = 84.7), and no interaction between the two factors, $F(6, 288) = 0.31$, $\eta_p^2 = .01$, Bayes Factor = 1123 in favor of the null. The second ANOVA was conducted using only those items which immediately followed the manipulated consolidation position. In this analysis there was a main effect of Consolidation Time, $F(2, 96) = 23.50$, $\eta_p^2 = .33$, Bayes Factor $= 1.2 * 10^7$ in favor of an effect (means: 217 ms = 84.7, 400 ms = 80.2, 700 ms = 72.7), a main effect of Serial Position, $F(2, 96) = 11.68$, $\eta_p^2 = .20$, Bayes Factor = 1,603 in favor of an effect (means: SP2 = 74.9, SP3 = 78.7, SP4 = 84.1), and no interaction between the two factors, $F(4, 192) = 2.16$, $\eta_p^2 = .04$, Bayes Factor = 10 in favor of the null.[2]

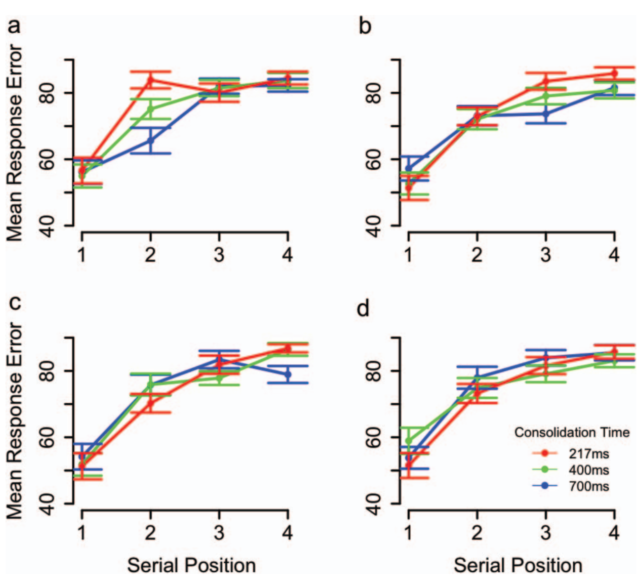

Figure 8. Mean response error in Experiment 4 as a function of Consolidation Time and Serial Position when the following Consolidation Periods were manipulated: (a) the first consolidation period, (b) the second consolidation period, (c) the third consolidation period, (d) the fourth consolidation period. Error bars represent standard error of the mean. See the online article for the color version of this figure.

Model comparison. The analysis of mean error shows that a consolidation effect is present only for the items immediately following the manipulated consolidation period. In order to understand this effect we perform our model-based analyses with only the data from items following the manipulated consolidation period. Table 5 lists model fit, as indexed by WAIC, for all of the models we tested using our stepwise procedure. The full model containing effects of consolidation time and serial position on all target parameters was the winner beating the next best-fitting model by 20 points. Table 2 lists all estimates of the target parameters under the full model as a function of consolidation time and serial position.

Bayes factor analysis of parameter values. Given that the best model selected during model comparison was the full model, we tested main effects and interactions for all factors and parameters. In brief, there was clear evidence of a main effect of Consolidation Time and Serial Position on P^M, the probability an item was in memory, and a main effect of Serial Position on P^O, the probability that a memory was of the specific-angle presented, but all other tests were either in favor of the null or ambiguous. The parameter means for the shortest and longest Consolidation Times can be found in Table 2.

For the P^M parameter, we found a main effect of Consolidation Time, Bayes factor $= 1.4 * 10^4$ in favor of an effect, and a main effect of Serial Position, Bayes factor $= 2.1 * 10^4$ in favor of an effect. There was no interaction, Bayes factor = 8.1 against an interaction. The pattern is that P^M decreases as Serial Position increases and increases as Consolidation Time increases.

For the imprecision parameter, σ^O, there was ambiguous evidence related to a main effect of Consolidation Time, Bayes factor = 1.3 in favor of an effect. There was nearly a meaningful effect of Serial Position, Bayes factor = 2.8 in favor of an effect. There was also ambiguous evidence against an interaction, Bayes factor = 2.0 against an interaction.

[2] Performance in this experiment was very low at some serial positions. To be sure our pattern of results was not influenced by floor effects we reanalyzed our basic ANOVA analyses after excluding all participants who performed at chance level for one or more serial positions (13 participants dropped, new $N = 36$). We found the same pattern of results, but with slightly stronger statistical findings.

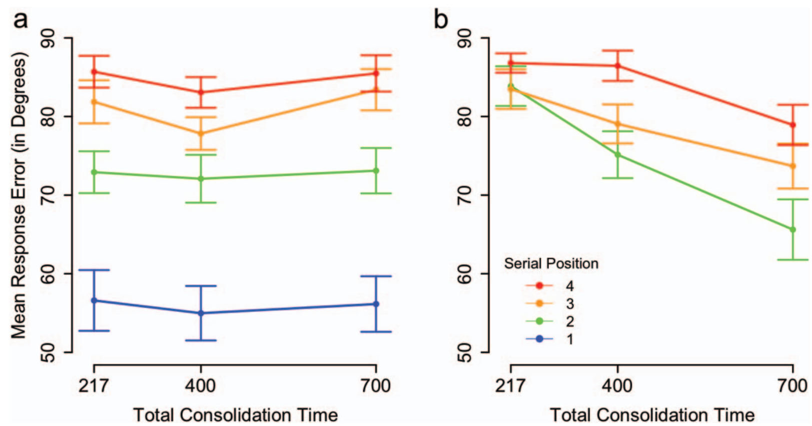

Figure 9. Mean response error in Experiment 4 as a function of Consolidation Time and Serial Position. (a) shows mean performance for serial positions immediately preceding the manipulated consolidation durations. (b) shows mean performance for serial positions immediately following the manipulated consolidation durations. Error bars represent standard error of the mean. See the online article for the color version of this figure.

Finally, for the P^O parameter, we found no main effect of Consolidation Time, Bayes factor = 3.4 against an effect. We found a main effect of Serial Position, Bayes factor = 7.3 in favor of an effect. There was ambiguous evidence that suggested that there was no interaction, with Bayes factor = 2.0 against an interaction. The pattern in the parameter means is complex and suggestive of an interaction, but the only interpretable effect is the main effect of Serial Position, which is essentially that P^O is higher at Serial Position 4.

Discussion

In Experiment 4 we manipulated only a single consolidation period on each trial. Under these conditions performance was affected for only the memory item immediately following the manipulated consolidation period. No other items were affected. This matches the prediction of attentional blink models of working memory consolidation (Lagroix et al., 2012; Wyble et al., 2009, 2011). Once consolidation of a memory item begins attention is no longer able to apprehend new items until the consolidation process is complete. This is in conflict with the theory that attention is free to switch to new items when they are presented. It is also in conflict with models that suggest free time between item presentation is used to cycle attention between memory items as a nonverbal maintenance mechanism (Barrouillet et al., 2013; Barrouillet & Camos, 2012), at least when the time between items is very brief.

Table 5
WAIC for All Tested Models in Experiment 4

Model	WAIC	Difference from best model
Full model	302,589	0
Constant P^M across serial position	302,710	121
Constant P^M across consolidation time	302,822	233
Constant σ^O across serial position	302,608	20
Constant σ^O across consolidation time	302,629	40
Constant P^O across serial position	302,711	122
Constant P^O across consolidation time	302,730	141

Note. WAIC = Watanabe-Akaike Information Criterion. These models were fit only on data from serial positions immediately following the manipulated consolidation period. In the full model parameters were estimated freely for all Serial Positions and Consolidation Times. In our stepwise procedure we estimated the full model but constrained a single parameter value to be invariant across Serial Position or Consolidation Time. WAIC and difference scores are rounded to the nearest whole number. P^M = the probability an item is in memory; σ^O = the standard deviation of the specific-angle response; P^O = the probability that a maintained memory is of the specific angle presented and not categorical in nature.

Although temporal distinctiveness is already ruled out as a cause of the short-term consolidation effect by Experiment 3, the present experiment provides further evidence against a temporal distinctiveness explanation. Under the prototypical temporal distinctiveness approach (i.e., Brown et al., 2007; Crowder, 1976), manipulation of a single consolidation period should not affect only the item immediately following the consolidation period. Rather, all items within the list should be affected. In particular the item immediately preceding the manipulated duration should be strongly affected. This is not what we observed. There was no effect on any other item in the list except for the item immediately following the manipulated consolidation period.

Alternative models of temporal distinctiveness that are based on the idea that trials are stored as events also have problems with the present results. In these approaches it is the events (individual trials) that are retrieved at recall so what is important is the distinctiveness of individual events (trials) from one another (Shipstead & Engle, 2013). When trials take longer to perform, such as when the consolidation periods are longer, then the trials become more distinct from one another and performance should improve, barring other factors that could influence distinctiveness.

The problem with this event-based distinctiveness approach is that if the individual events become more distinct with larger consolidation durations then manipulating a single consolidation period should affect performance for all items on the trial. This is because distinctiveness works on the trial level not the individual item level. This was not observed in our data. Further, the change in distinctiveness in our manipulation was extremely small, yet still produced a sizable consolidation effect. This is difficult to explain from an event-based model. In related work we have also shown that event-based temporal distinctiveness does not function in visual working memory change detection tasks (Ricker et al., 2014). In all, there is strong evidence against any form of distinctiveness in the present work.

Our modeling analysis again provided evidence that the driver of the consolidation effect is the probability that an item is in memory at test, P^M. We found evidence against a qualitative change in the nature of the memory maintained, P^O, with changes to consolidation time, and ambiguous evidence for a change in the imprecision of the specific-angle maintained, σ^O, with longer consolidation times. The ambiguity in evidence for an effect of consolidation time on σ^O may be an artifact of the low level of performance in some serial positions of this experiment. With a small number of responses based upon the specific angle presented the model will have difficulty estimating σ^O, resulting in ambiguous evidence for or against an effect. In all, the findings of Experiment 4 are in agreement with those of Experiment 2 and in partial agreement with Experiment 1.

General Discussion

In the present work we examine the establishment of a stable visual working memory representation. Although the topic of short-term consolidation is not a new one, previous work has not provided clarity about how increased processing during memory consolidation improves memory performance. Uncertainty over what exactly occurs during consolidation has even led to debate over whether masking ends the consolidation process (Jolicœur & Dell'Acqua, 1998; Vogel et al., 2006). Elsewhere, we and others have shown that masking does not end the consolidation process within working memory (Nieuwenstein & Wyble, 2014; Ricker & Cowan, 2014) and we again confirm this finding. The present work also establishes that the consolidation effect in visual working memory is not restricted to recognition paradigms, but is also observed when memory recall is required.

The main goal of this study was to determine what process leads to the consolidation effect in working memory. We began by applying the mathematical model of Hardman et al. (2017) to our delayed estimation data to determine how consolidation changes memory for visual items. This model is elucidating in that it assumes participants sometimes have memory for the specific angle where the stimulus was presented and other times only memory of the general categorical location in which the stimulus was presented, such as "left side." By estimating these categorical contributions to memory, we were able to determine whether consolidation of memory affects one or more of the following processes: (1) the probability that an item will be held in memory at the time of test, (2) the precision of memory for the specific angle presented, or (3) the probability that items held in memory will contain the specific angle presented rather than only the category to which the presented item belonged.

In the introduction, we discussed four possible mechanisms that could explain why changing consolidation time could affect performance. The first, stabilization of representational precision, predicted that memories should become more precise with additional consolidation time. The second mechanism, changing the nature of stabilized representations, predicts that items in WM should be more likely to be held with specific-angle rather than categorical information. The third and fourth mechanisms, the attentional blink and temporal distinctiveness, respectively, predict that items should be more likely to be in WM given more consolidation time. These last two mechanisms differ in predictions about how this change should be distributed across time and serial positions.

Our model-based analysis shows that the consolidation effect is primarily due to changes in (1) the probability that an item will be held in memory at the time of test. This effect is large and consistent across all three experiments that demonstrate a consolidation effect. In Experiment 1 we also observed changes in (2) the precision of specific-angle memory and (3) qualitative shifts in the nature of the memory maintained with changes to consolidation time, which are predicted by the first and second mechanisms, respectively. These effects, however, were smaller in magnitude and did not replicate in Experiments 2 and 4. It should be noted that in Experiment 4 our estimate of the specific-angle precision may have been ambiguous because of a relatively small number of responses contributing to its estimation. On balance it seems that if there are changes to the precision or qualitative nature of memory with longer consolidation times, these changes account for a minority of the observed effect or are not consistently present. Thus, we reject the first two mechanisms as being clearly relevant.

The strong and consistent effect we found was that consolidation time affects the probability that a memory item will be available at test. There are two relevant theories that could explain this effect. The attentional blink approach describes the change as a failure of attention to orient and consolidate all memory items with short consolidation durations (Lagroix et al., 2012; Nieuwenstein, Potter, et al., 2009; Wyble et al., 2009, 2011). The temporal distinctiveness approach describes this change as a function of retrieval cue utility (Brown et al., 2007; Crowder, 1976; Souza & Oberauer, 2015; Shipstead & Engle, 2013). We provide strong evidence that temporal distinctiveness is not at work here and that the consolidation effect in working memory is attributable to an attentional blink.

In Experiment 3, we used consolidation times longer than the hypothesized consolidation interval and varied those consolidation times between trials. Temporal distinctiveness theories predict that greater separation between items on trials with longer consolidation intervals would result in more temporally distinct and, therefore, easier to recall representations. Instead, we found no effect of consolidation time. This result, however, fits the attentional blink hypothesis, which predicts no effect of consolidation time beyond an initial critical period. Additionally, in Experiment 4, we manipulated a single consolidation interval within a trial. Temporal distinctiveness theories would predict that this would affect all of the items on the trial, to greater or lesser extents, depending on their proximity to the manipulated interval. Instead, we only found an effect on the item following the manipulated consolidation interval. Again, this result fits the attentional blink hypothesis, which predicts that items presented during the consolidation of a previous item are likely to be missed during the attentional blink.

In studies on the attentional blink a string of stimuli, often digits, are presented rapidly at a pace of about 100 ms per item with two targets, often letters, placed somewhere within the stream. The task of participants is to correctly identify both letters. Performance is generally quite good. Critically, accuracy for the second target suffers when it is presented within 200 – 500 ms after the first target, although some authors put the outer limit of impairment as far as 700 ms (Lagroix et al., 2012). Recent evidence strongly indicates that this effect is due to attention not being available to consolidate the second item while consolidation of the first item is ongoing (Nieuwenstein, Potter, et al., 2009; Wyble et al., 2009, 2011).

Similarly, in our experiments we find that performance for the first item presented is always quite good irrespective of the time available for consolidation. Longer consolidation times are beneficial for all other items on the list. This sparing for serial position 1 suggests that for consolidation of a subsequent item to fail, an attentional blink must be initiated by consolidation of the previous item. Experiment 4 confirms this intuition, demonstrating that altering consolidation period durations only affects performance for the item immediately following the manipulated duration.

We propose that the consolidation effect is driven by failure to create stable working memory traces for all presented items. This occurs when a new item is presented during the attentional blink of the previous item. Attention becomes stuck on the preceding item during the blink and is only able to switch to the next item after significant degradation of the newly presented and still unstable sensory memory trace. If there is a smaller effect of consolidation time on the precision and the relative-probability of a specific-angle memory these likely are attributable to attention sometimes shifting to a degraded, but still present, sensory trace. These traces may carry less information as the time available for consolidation shortens due to trace decay of the fine-details they contain or interference from the masking stimuli. Faster consolidation should preserve a higher-precision sensory memory and/or the presence of the specific-angle representation.

Tentative Theories in Need of Future Investigation

Evidence that the consolidation process stabilizes the memory trace and reduces future forgetting is provided by Ricker and Cowan (2014). In that work, conditions which contain more time for consolidation of the memory trace result in slower rates of time-based forgetting across a later retention interval. The present work also shows reduced forgetting under conditions with more consolidation time. In each experiment conditions with more consolidation time show a slower loss in performance as the serial position increases. In other words, the difference in performance between serial position 1 and 4 is greater with less time for consolidation.

Presentation order and output order were confounded in our procedure so we cannot attribute the locus of this effect to a reduction in presentation or output effects specifically. It is plausible that consolidation of the memory trace reduced output interference during the response period, but a reduction in serial order effects could also be playing a role. The potential role of reduced output interference is more consistent with our stabilization theory of consolidation in which short-term consolidation reduces later forgetting of the memory trace. Future work should examine whether output interference specifically, and interference more generally, is reduced by increasing the time available for consolidation.

A small but interesting effect we see most clearly in Experiment 2, and to some degree in Experiments 1 and 4, is that longer consolidation times hurt performance for the first item in the list. Although the first item in the list is always fully consolidated there is a small level of baseline forgetting that continues to occur. Our present work is not designed to investigate why this occurs, but we do have some post hoc explanations of what is happening.

With the very brief periods between item presentations in Experiments 1, 2, and 4 there is no time to consolidate an item then switch cognitive tasks to active maintenance. Items already in mind are not actively maintained during the presentation sequence and suffer some time-based forgetting (Ricker & Cowan, 2010; Ricker, Vergauwe, & Cowan, 2016). In the case of serial positions 2–4, the amount of decay that they suffer is not observable because the benefit from a longer consolidation period is large while the decay that occurs is quite small. For serial position 1 there is no benefit to longer consolidation times, so the memory decay over time is observed. We do not observe this same effect in Experiment 3 because there is enough time between item presentations to fully consolidate each item then switch to active maintenance of other items in the list, countering memory decay.

Relationship to Attention and Retrieval Theories of Visual Working Memory

Previous work arguing that the attentional blink is a result of ongoing memory consolidation has been based upon studies using rapid serial visual presentation (RSVP) of stimuli (Lagroix et al., 2012; Nieuwenstein, Potter, et al., 2009; Nieuwenstein, Van der Burg, et al., 2009; Wyble et al., 2009, 2011). The RSVP paradigm is quite different from a standard visual working memory task in that RSVP involves the presentation of 10 or more stimuli a second. This creates a large amount of visual interference involving strong forward and backward masking. It would be reasonable to assume that this high level of perceptual demand is necessary to observe an attentional blink and that no blink would be observed in a standard visual working memory task. The present work shows that this is not the case. Here we show evidence for a common consolidation mechanism across these very different experimental conditions.

The present study also provides a strong counterargument to temporal distinctiveness models of visual working memory performance. Temporal distinctiveness approaches to visual memory (Souza & Oberauer, 2015; Shipstead & Engle, 2013) account for improvement in memory performance with more time for consolidation between item presentations by arguing that memory is encoded along a temporal dimension. When retrieving memories at test, the temporal dimension is used as a retrieval cue and items that have similar temporal values will interfere with one another for expression at retrieval, according to these models.

Although these models may be able to provide a good fit for our data in Experiments 1 and 2, the lack of any temporal spacing effect in Experiment 3 is problematic. Even with conservative scaling of the temporal dimension, we still see predictions of robust effects from the temporal distinctiveness calculations following the prominent model of Brown et al. (2007; see Table 4). To our knowledge all authors of temporal distinctiveness accounts of visual working memory performance describe their theoretical approach as consistent with the Brown et al. model that we disprove here. Experiment 4 provides more evidence against temporal distinctiveness by showing strong effects of temporal spacing that are isolated to only the item immediately following the manipulated consolidation period. These spacing effects should modify the distinctiveness of all memory items in the trial, in conflict with what was observed.

This inability of temporal distinctiveness approaches to account for the observed temporal spacing effects is in agreement with Ricker et al. (2014) who show a similar finding with change detection for unfamiliar character arrays. In that study the temporal spacing between trials was manipulated but performance did not change in two experiments. In a further two experiments the change in performance that followed from the temporal spacing manipulation did not follow the predictions of the temporal distinctiveness approach. Ricker and Cowan (2014) use the same change detection procedure with the same stimulus set as Ricker et al. and found an effect of consolidation time on visual working memory performance. The convergence of results across paradigms and stimuli calls into question the viability of temporal distinctiveness approaches within visual working memory more generally.

Concluding Remarks

The importance of our method for unveiling the key findings presented here cannot be overstated. Although it is common to aggregate data across many variables, such as serial position and presented stimulus orientation, aggregating over serial position would have masked the lack of an effect of consolidation at the first serial position. The interaction across serial positions in our present work gives critical evidence that working memory consolidation and the attentional blink are the same process. Lack of aggregation of our data across presented stimulus orientations allowed us to differentiate between the changing quality of working memory representations and whether or not items were in memory at the time of test. The importance of this second aspect for accurate model estimation is addressed in length by Hardman et al. (2017).

The connections between the attentional blink and visual working memory may be the most important contribution of the present work. Future research is needed to discover questions such as the role of masking in the memory creation process and whether attention is completely suppressed during short-term consolidation. There is now more reason than ever that research from both fields should be used to inform and drive theory in unison, bringing us closer to a unified theory of attention, perception, and memory.

References

Asplund, C. L., Fougnie, D., Zughni, S., Martin, J. W., & Marois, R. (2014). The attentional blink reveals the probabilistic nature of discrete conscious perception. *Psychological Science, 25,* 824–831. http://dx.doi.org/10.1177/0956797613513810

Bae, G.-Y., Olkkonen, M., Allred, S. R., & Flombaum, J. I. (2015). Why some colors appear more memorable than others: A model combining categories and particulars in color working memory. *Journal of Experimental Psychology: General, 144,* 744–763. http://dx.doi.org/10.1037/xge0000076

Baddeley, A. (2000). The episodic buffer: A new component of working memory? *Trends in Cognitive Sciences, 4,* 417–423. http://dx.doi.org/10.1016/S1364-6613(00)01538-2

Barrouillet, P., & Camos, V. (2012). As time goes by temporal constraints in working memory. *Current Directions in Psychological Science, 21,* 413–419. http://dx.doi.org/10.1177/0963721412459513

Barrouillet, P., Plancher, G., Guida, A., & Camos, V. (2013). Forgetting at short term: When do event-based interference and temporal factors have an effect? *Acta Psychologica, 142,* 155–167. http://dx.doi.org/10.1016/j.actpsy.2012.12.003

Bayliss, D. M., Bogdanovs, J., & Jarrold, C. (2015). Consolidating working memory: Distinguishing the effects of consolidation, rehearsal and attentional refreshing in a working memory span task. *Journal of Memory and Language, 81,* 34–50. http://dx.doi.org/10.1016/j.jml.2014.12.004

Bays, P. M., Gorgoraptis, N., Wee, N., Marshall, L., & Husain, M. (2011). Temporal dynamics of encoding, storage, and reallocation of visual working memory. *Journal of Vision, 11,* 6. http://dx.doi.org/10.1167/11.10.6

Brown, G. D. A., Neath, I., & Chater, N. (2007). A temporal ratio model of memory. *Psychological Review, 114,* 539–576. http://dx.doi.org/10.1037/0033-295X.114.3.539

Chun, M. M., & Potter, M. C. (1995). A two-stage model for multiple target detection in rapid serial visual presentation. *Journal of Experimental Psychology: Human Perception and Performance, 21,* 109–127. http://dx.doi.org/10.1037/0096-1523.21.1.109

Cleveland, W. S., Grosse, E., & Shyu, W. M. (1992). Local regression models. In J. M. Chambers & T. J. Hastie (Eds.), *Statistical models in S* (pp. 309–376). Pacific Grove, CA: Wadsworth & Brooks.

Cowan, N. (1984). On short and long auditory stores. *Psychological Bulletin, 96,* 341–370. http://dx.doi.org/10.1037/0033-2909.96.2.341

Cowan, N. (1988). Evolving conceptions of memory storage, selective attention, and their mutual constraints within the human information-processing system. *Psychological Bulletin, 104,* 163–191. http://dx.doi.org/10.1037/0033-2909.104.2.163

Cowan, N. (1995). *Attention and memory: An integrated framework*. Oxford, England: Oxford University Press.

Crowder, R. G. (1976). *Principles of learning and memory*. Hillsdale, NJ: Erlbaum.

Donkin, C., Nosofsky, R., Gold, J., & Shiffrin, R. (2015). Verbal labeling, gradual decay, and sudden death in visual short-term memory. *Psychonomic Bulletin & Review, 22,* 170–178. http://dx.doi.org/10.3758/s13423-014-0675-5

Edwards, W., Lindman, H., & Savage, L. J. (1963). Bayesian statistical inference for psychological research. *Psychological Review, 70,* 193–242. http://dx.doi.org/10.1037/h0044139

Fougnie, D., Suchow, J. W., & Alvarez, G. A. (2012). Variability in the quality of visual working memory. *Nature Communications, 3,* 1229. http://dx.doi.org/10.1038/ncomms2237

Gelman, A., Hwang, J., & Vehtari, A. (2014). Understanding predictive information criteria for Bayesian models. *Statistics and Computing, 24,* 997–1016. http://dx.doi.org/10.1007/s11222-013-9416-2

Goodman, S. N. (1999). Toward evidence-based medical statistics. 1: The P value fallacy. *Annals of Internal Medicine, 130,* 995–1004. http://dx.doi.org/10.7326/0003-4819-130-12-199906150-00008

Hardman, K. O. (2016). CatContModel: Categorical and continuous working memory models for delayed estimation tasks (Version 0.7.0) [Computer software]. Retrieved from https://github.com/hardmanko/CatContModel/releases/tag/v0.7.0.

Hardman, K. O., Vergauwe, E., & Ricker, T. J. (2017). Categorical working memory representations are used in delayed estimation of continuous colors. *Journal of Experimental Psychology: Human Perception and Performance, 43,* 30–54. http://dx.doi.org/10.1037/xhp0000290

Jolicœur, P., & Dell'Acqua, R. (1998). The demonstration of short-term consolidation. *Cognitive Psychology, 36,* 138–202. http://dx.doi.org/10.1006/cogp.1998.0684

Kass, R. E., & Raftery, A. E. (1995). Bayes factors. *Journal of the American Statistical Association, 90,* 773–795. http://dx.doi.org/10.1080/01621459.1995.10476572

Kooperberg, C. (2015). polspline: Polynomial spline routines (Version 1.1.12) [Computer software]. Retrieved from https://CRAN.R-project.org/package=polspline

Kutner, M. H., Nachtsheim, C. J., Neter, J., & Li, W. (2004). *Applied Linear Statistical Models* (5th ed.). New York, NY: McGraw-Hill/Irwin.

Lagroix, H. E., Spalek, T. M., Wyble, B., Jannati, A., & Di Lollo, V. (2012). The root cause of the attentional blink: First-target processing or disruption of input control? *Attention, Perception, & Psychophysics, 74,* 1606–1622. http://dx.doi.org/10.3758/s13414-012-0361-5

Ma, W. J., Husain, M., & Bays, P. M. (2014). Changing concepts of working memory. *Nature Neuroscience, 17,* 347–356. http://dx.doi.org/10.1038/nn.3655

Massaro, D. W. (1975). Backward recognition masking. *The Journal of the Acoustical Society of America, 58,* 1059–1065. http://dx.doi.org/10.1121/1.380765

Morey, R. D., & Rouder, J. N. (2012). BayesFactor: Computation of Bayes factors for simple designs (R package Version 0.9.12–2) [Computer software]. Retrieved from http://CRAN.R-project.org/package=BayesFactor

Nieuwenstein, M. R., Potter, M. C., & Theeuwes, J. (2009). Unmasking the attentional blink. *Journal of Experimental Psychology: Human Perception and Performance, 35,* 159–169. http://dx.doi.org/10.1037/0096-1523.35.1.159

Nieuwenstein, M., Van der Burg, E., Theeuwes, J., Wyble, B., & Potter, M. (2009). Temporal constraints on conscious vision: On the ubiquitous nature of the attentional blink. *Journal of Vision, 9,* 1–14. http://dx.doi.org/10.1167/9.9.18

Nieuwenstein, M., & Wyble, B. (2014). Beyond a mask and against the bottleneck: Retroactive dual-task interference during working memory consolidation of a masked visual target. *Journal of Experimental Psychology: General, 143,* 1409–1427. http://dx.doi.org/10.1037/a0035257

Oberauer, K., Lewandowsky, S., Farrell, S., Jarrold, C., & Greaves, M. (2012). Modeling working memory: An interference model of complex span. *Psychonomic Bulletin & Review, 19,* 779–819. http://dx.doi.org/10.3758/s13423-012-0272-4

Ouimet, C., & Jolicoeur, P. (2007). Beyond Task 1 difficulty: The duration of T1 encoding modulates the attentional blink. *Visual Cognition, 15,* 290–304. http://dx.doi.org/10.1080/13506280600693741

Raymond, J. E., Shapiro, K. L., & Arnell, K. M. (1992). Temporary suppression of visual processing in an RSVP task: An attentional blink? *Journal of Experimental Psychology: Human Perception and Performance, 18,* 849–860. http://dx.doi.org/10.1037/0096-1523.18.3.849

R Core Team. (2015). R: A language and environment for statistical computing [Computer software manual]. Vienna, Austria. Retrieved from https://www.R-project.org/

Ricker, T. J. (2015). The role of short-term consolidation in memory persistence. *AIMS Neuroscience, 2,* 259–279. http://dx.doi.org/10.3934/Neuroscience.2015.4.259

Ricker, T. J., & Cowan, N. (2010). Loss of visual working memory within seconds: The combined use of refreshable and non-refreshable features. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 36,* 1355–1368. http://dx.doi.org/10.1037/a0020356

Ricker, T. J., & Cowan, N. (2014). Differences between presentation methods in working memory procedures: A matter of working memory consolidation. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 40,* 417–428. http://dx.doi.org/10.1037/a0034301

Ricker, T. J., Spiegel, L. R., & Cowan, N. (2014). Time-based loss in visual short-term memory is from trace decay, not temporal distinctiveness. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 40,* 1510–1523. http://dx.doi.org/10.1037/xlm0000018

Ricker, T. J., Vergauwe, E., & Cowan, N. (2016). Decay theory of immediate memory: From Brown (1958) to today (2014). *Quarterly Journal of Experimental Psychology, Published Online*. http://dx.doi.org/10.1080/17470218.2014.914546

Rouder, J. N., Morey, R. D., Speckman, P. L., & Province, J. M. (2012). Default Bayes factors for ANOVA designs. *Journal of Mathematical Psychology, 56,* 356–374. http://dx.doi.org/10.1016/j.jmp.2012.08.001

Rouder, J. N., Morey, R. D., Verhagen, J., Province, J. M., & Wagenmakers, E.-J. (2016). Is there a free lunch in inference? *Topics in Cognitive Science, 8,* 520–547. http://dx.doi.org/10.1111/tops.12214

Rouder, J. N., Speckman, P. L., Sun, D., Morey, R. D., & Iverson, G. (2009). Bayesian t tests for accepting and rejecting the null hypothesis. *Psychonomic Bulletin & Review, 16,* 225–237. http://dx.doi.org/10.3758/PBR.16.2.225

Shipstead, Z., & Engle, R. W. (2013). Interference within the focus of attention: Working memory tasks reflect more than temporary maintenance. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 39,* 277–289. http://dx.doi.org/10.1037/a0028467

Souza, A. S., & Oberauer, K. (2015). Time-based forgetting in visual working memory reflects temporal distinctiveness, not decay. *Psychonomic Bulletin & Review, 22,* 156–162. http://dx.doi.org/10.3758/s13423-014-0652-z

Sperling, G. (1960). The information available in brief visual presentations. *Psychological Monographs, 74,* 1–29. http://dx.doi.org/10.1037/h0093759

Stevanovski, B., & Jolicœur, P. (2007). Visual short-term memory: Central capacity limitations in short-term consolidation. *Visual Cognition, 15,* 532–563. http://dx.doi.org/10.1080/13506280600871917

Sy, J., Marois, R., & Tong, F. (2015). Degraded precision of consciously perceived targets in the attentional blink. *Journal of Vision, 15,* 1234. http://dx.doi.org/10.1167/15.12.1234

van den Berg, R., Awh, E., & Ma, W. J. (2014). Factorial comparison of working memory models. *Psychological Review, 121,* 124–149. http://dx.doi.org/10.1037/a0035234

Vogel, E. K., Woodman, G. F., & Luck, S. J. (2006). The time course of consolidation in visual working memory. *Journal of Experimental Psychology: Human Perception and Performance, 32,* 1436–1451. http://dx.doi.org/10.1037/0096-1523.32.6.1436

Wagenmakers, E. J. (2007). A practical solution to the pervasive problems of p values. *Psychonomic Bulletin & Review, 14,* 779–804. http://dx.doi.org/10.3758/BF03194105

Wagenmakers, E.-J., Lodewyckx, T., Kuriyal, H., & Grasman, R. (2010). Bayesian hypothesis testing for psychologists: A tutorial on the Savage-Dickey method. *Cognitive Psychology, 60,* 158–189. http://dx.doi.org/10.1016/j.cogpsych.2009.12.001

Wagenmakers, E. J., Wetzels, R., Borsboom, D., & van der Maas, H. L. (2011). Why psychologists must change the way they analyze their data: The case of psi: Comment on Bem (2011). *Journal of Personality and Social Psychology, 100,* 426–432. http://dx.doi.org/10.1037/a0022790

Wilken, P., & Ma, W. J. (2004). A detection theory account of change detection. *Journal of Vision, 4,* 1120–1135. http://dx.doi.org/10.1167/4.12.11

Woodman, G. F., & Vogel, E. K. (2005). Fractionating working memory: Consolidation and maintenance are independent processes. *Psychological Science, 16,* 106–113. http://dx.doi.org/10.1111/j.0956-7976.2005.00790.x

Woodman, G. F., & Vogel, E. K. (2008). Selective storage and maintenance of an object's features in visual working memory. *Psychonomic Bulletin & Review, 15,* 223–229. http://dx.doi.org/10.3758/PBR.15.1.223

Wyble, B., Bowman, H., & Nieuwenstein, M. (2009). The attentional blink provides episodic distinctiveness: Sparing at a cost. *Journal of Experimental Psychology: Human Perception and Performance, 35,* 787–807. http://dx.doi.org/10.1037/a0013902

Wyble, B., Potter, M. C., Bowman, H., & Nieuwenstein, M. (2011). Attentional episodes in visual perception. *Journal of Experimental Psychology: General, 140,* 488–505. http://dx.doi.org/10.1037/a0023612

Zhang, W., & Luck, S. J. (2008). Discrete fixed-resolution representations in visual working memory. *Nature, 453,* 233–235. http://dx.doi.org/10.1038/nature06860

(*Appendix follows*)

Appendix

Hypothesis Tests on Main Effects and Interactions

For the analyses of the parameters estimated from the Hardman et al. (2017) model, we performed some Bayesian hypothesis tests on main effects and interactions between experimental factors, such as serial position and consolidation time. The technique that we used has not, in our knowledge, been published, was not straightforward and requires some explanation, which is given here. The technique is fairly general and can be applied to a range of Bayesian models. It is conceptually equivalent to analysis of variance (ANOVA), although the mechanics of the technique are fairly different from ANOVA. The technique only works for fully crossed designs, where you have data in every cell of the design. Note that although we used a within-participants design, this hypothesis testing technique does not require a within-participants design, only that the model has parameters that account for differences between conditions. The reader may find it helpful to read the Appendix of Hardman et al. (2017) for information on the specifics of the specification of that model.

In the Hardman et al. (2017) model, parameter values are allowed to vary between task conditions. For example, the P^M parameter varies between conditions. This is done by estimating a single parameter that is shared by all participants that is added to all participant-level parameters, shifting each participant's parameter by the same amount. For example, each participant has their own P^M parameter that reflects their baseline performance, but their effective performance in any condition is based on their own P^M parameter plus the effect of task condition. We call the single, shared parameter for each condition a *condition effect*.

The condition effects are differences from a cornerstone condition. Other models may instead estimate cell means or differences from a grand mean. For the purposes of this technique, there is no substantial difference between true cell means, differences from a grand mean, or differences from a cornerstone condition. It is conceptually most simple to think of the condition effects as cell means, so we will do so.

This hypothesis testing technique is for fully crossed factorial experimental designs and it can be thought of as a kind of ANOVA. This hypothesis testing technique works for experimental designs with any number of factors, but we will use a two-factor design as an example. The example will be a design with two levels of the A factor and three levels of the B factor. You could think of the A factor as consolidation time and the B factor as serial position. A two-factor ANOVA model can be written as

$$\mu_{ij} = \mu_G + \alpha_i + \beta_j + \pi_{ij}$$

where μ_{ij} is the cell mean (or, in our case, condition effect) for the ith row and jth column of the design, μ_G is the grand mean, α_i is the main effect parameter for the ith row, β_j is the main effect parameter for the jth column, and π_{ij} is the interaction parameter for the ith row and jth column. The main effect and interaction parameters are constructed with sums-to-zero constraints on the rows and columns. This means that the αs and βs sum to zero, as do the πs within each row or column. For the sake of clarity, we will use the term *ANOVA parameters* for μ_G, α_i, β_j, and π_{ij}.

It is possible to construct a design matrix, X, that satisfies the equality

$$\mu = X\theta$$

where μ is a vector of cell means (or condition effects) and θ is a vector of ANOVA parameters. Some of the ANOVA parameters are not included in θ because of the sums to zero constraints. We will choose to drop the first level of each of the factors. As a result, for the 2 by 3 example,

$$\theta = [\mu_G \alpha_2 \beta_2 \beta_3 \pi_{22} \pi_{23}]^T$$

where superscript T indicates transpose (i.e., θ is a column vector). The sums-to-zero constraints are enforced by the way in which X is constructed. This can be done in R (R Core Team, 2015) with the "model.matrix" function using the "contr.sum" function for creating sums-to-zero contrasts.

In this hypothesis testing technique, as in standard ANOVA, we have the cell means, μ, and must solve for θ. This is straightforward and results in the solution

$$\theta = (X^T X)^{-1} X^T \mu$$

For the derivation of this equation, see a GLM textbook (e.g., Kutner, Nachtsheim, Neter, & Li, 2004; Ch. 5, p. 199). We will define

$$S = (X^T X)^{-1} X^T$$

where S can be thought of as the Matrix that converts cell means into ANOVA parameters.

In the Hardman et al. (2017) model, each of the condition effect parameters has a prior on it. The priors are independent Cauchy distributions with location 0 and a scale that depends on which model parameter is being used. See Hardman et al. for more information on the default priors. The θ vector contains parameters calculated from the cell means, μ. Thus, the priors on the θs depend on the priors on the μs. The priors on the μs are 0 centered and the Cauchy distribution is symmetrical, which results in the marginal priors on the θs being 0 centered. It is not necessary to know more about the priors on the θs for the purpose of this technique.

For the sake of example, we will examine a test the main effect of the B factor. We would use the 2 parameters in θ that are related to the main effect of B: β_2 and β_3. There are three levels of B, but the sums to zero constraints mean that there are only two βs in θ. We will need β_1 as well, which we can calculate from β_2 and β_3 using the knowledge that the βs sum to 0. The implied β, β_1, is equal to the negative sum of the other betas: $\beta_1 = -(\beta_2 + \beta_3)$. When an interaction is being tested, calculating implied interaction

(*Appendix continues*)

terms is more complex, but simply requires that the same procedure be applied iteratively to rows and columns.

We can calculate the βs based on both the prior and posterior cell means separately. As discussed, the prior on each of the βs is 0 centered. As such, the prior belief is that the βs are dispersed around 0. In the posterior, the βs may be more or less dispersed around 0 than in the prior. If there is no main effect, we would expect the posterior βs to be more closely clustered around 0 than the prior βs. If there is a main effect, we would expect that the posterior βs would be more dispersed from 0 than in the prior. This logic establishes the basic setup for the hypothesis test, but there are a few more steps.

We will use the Savage-Dickey density ratio as part of this hypothesis testing technique. The Savage-Dickey density ratio can be used to provide the Bayes factor for a test of the hypothesis that a parameter has a given value (Wagenmakers, Lodewyckx, Kuriyal, & Grasman, 2010). It requires the estimation of the density at a point in the prior distribution and at the same point in the posterior distribution. Under the null hypothesis that there is no main effect of the B factor, we want to test that the β parameters are all at 0. Thus, we need to estimate the prior and posterior densities at **0** (the zero vector, of length equal to the number of β parameters). This is complicated by the fact that it is nontrivial to estimate the density of a multidimensional distribution. Thus, to use Savage-Dickey, we must reduce the dimensionality of the problem: Rather than working with the prior and posterior distributions of the βs, we would like to collapse the βs down to, ideally, 1 dimension.

The hypothesis we want to test relates to how near to 0 the βs are. Thus, some measure of the dispersion of the βs around 0 would be an appropriate way to reduce the dimensionality of the problem. We will call this measure of dispersion Δ. We chose to use the sample variance to measure dispersion, so we define $\Delta = \text{Var}(\beta)$. We chose to use variance because it resulted in relatively stable density estimates when compared with other measures of dispersion that we tried, such as standard deviation or the length of the β vector.

We are interested in testing the hypothesis that all βs are 0, but using Δ instead of the βs. If all of the βs are 0, then $\Delta = 0$. Thus, to calculate the Savage-Dickey density ratio, we should compare the prior and posterior densities of Δ at 0. We used the polspline package (Kooperberg, 2015) for R for kernel density estimation as suggested by Wagenmakers et al. (2010).

We are able to sample directly from the prior distribution of the cell means (μ), because the model places priors directly on the cell means. We are also able to indirectly sample from the posterior distribution of the cell means, which is what happens during parameter estimation. We do not, however, know how to sample directly from the prior or posterior distribution of the ANOVA parameters. However, given samples from the prior and posterior distributions of the μs and the matrix *S*, we can calculate the prior and posterior distributions of the ANOVA parameters with the equation $\theta = S\mu$. This calculation done for each sample from the prior or posterior of μ, resulting in prior and posterior distributions of θ.

To summarize, the steps of the procedure follow. 1. Sample from the posterior distribution of the cell means by doing model parameter estimation. 2. Sample from the prior distribution of the cell means. Take a number of samples equal to the number of samples taken from the posterior distribution, less burn-in iterations. 3. Create X and calculate S. Do the following steps separately for the prior and posterior distributions: 4a. Using S and the cell means, calculate θ (the ANOVA parameters). Do this for each sample of μ (i.e., each MCMC iteration or each sample from the prior). 4b. For each test of interest (i.e., main effects or interactions), select from θ the appropriate subset of ANOVA parameters (e.g., the βs for the test of the main effect of factor B). Do this for each sample of θ. 4c. For each sample of θ, calculate the ANOVA parameter(s) implied by the sums-to-zero constraint (e.g., β_1 from β_2 and β_3). 4d. Calculate Δ for each sample of ANOVA parameters (e.g., βs). 4e. Estimate the density of Δ at 0 with, for example, the polspline package. 5. Calculate the Savage-Dickey density ratio by dividing the prior density by the posterior density. This is an estimate of the Bayes factor in favor of the hypothesis that there is an effect (a main effect or interaction, depending on what is being tested). All of the preceding steps must be done separately for each model parameter (e.g., P^M).

To obtain an estimate of the variability in Bayes factors, you can repeat this procedure either with 1. subsets of the iterations that form the posterior distribution and/or 2. with different samples from the prior. Taking new samples from the prior is very computationally cheap, so we used that approach. This whole procedure, including estimating the variability in the Bayes factors, is a part of the CatContModel package (Hardman, 2016) and can be accessed with the "testMainEffectsAndInteractions" function.

Received August 11, 2016
Revision received May 26, 2017
Accepted May 31, 2017 ■