Journal of Experimental Psychology:
Human Perception and Performance
2014, Vol. 40, No. 3, 1237–1256


0096-1523/14/$12.00 DOI: 10.1037/a0036331

# Unloading and Reloading Working Memory: Attending to One Item Frees Capacity

Alessandra S. Souza, Laura Rerko, and Klaus Oberauer
University of Zurich

During the retention interval of a working memory task, presenting a retro-cue directs attention to 1 of the items in working memory. Testing the cued item leads to faster and more accurate responses. We contrasted 5 explanations of this benefit: (a) removal of noncued items, (b) strengthening of the cued item, (c) protection from probe interference, (d) protection from degradation, and (e) prioritization during the decision process. Experiment 1 showed that retro-cues reduced the set size effect in a visual recognition task, and did so increasingly with more time available to use the retro-cue. This finding is predicted only by Hypotheses 1 and 2. Hypotheses 3 through 5 were ruled out as explanations of the retro-cue benefit in this experiment. In Experiments 2 and 3, participants encoded 2 sequentially presented memory sets. In half of the trials, 1 item from the first set was retro-cued during the interset interval. Retro-cues improved memory for the second set. This reloading benefit is predicted only by the removal hypothesis: Irrelevant contents are removed from working memory, freeing capacity to encode new contents. Experiment 3 also yielded evidence that strengthening of the cued item might contribute to the retro-cue effect.

*Keywords:* working memory, attention, removal, retro-cue, set size

*Supplemental materials:* http://dx.doi.org/10.1037/a0036331.supp



Working memory (WM) is the system responsible for maintaining, in an accessible state, the subset of information that is needed to perform a given cognitive task. The capacity of WM is limited[1] (e.g., Cowan, 2010; Luck & Vogel, 1997). This limitation is evident, for example, in the set size effect: When the number of items to be maintained in WM is increased, performance (i.e., reaction time [RT] and accuracy) decreases in a monotonic fashion (e.g., Luck & Vogel, 1997; Sternberg, 1966, 1975).

Despite its capacity limitation, WM contents are managed flexibly: The current contents of WM are assumed to be the ones most relevant to the task at hand. This implies a fine-tuned modulation of what is maintained in WM as relevance changes over time. Selective attention has been considered one of the likely mechanisms supporting the flexible foregrounding of information in WM (see Gazzaley & Nobre, 2012, for a review). The effects of selective attention on information in WM have been investigated using the retro-cue paradigm (Griffin & Nobre, 2003; Landman, Spekreijse, & Lamme, 2003). In this paradigm, a retro-cue is presented during the retention interval of a WM recognition task, and this cue indicates the most relevant item for a subsequent memory test. Testing of the cued item leads to faster and more accurate responses compared with testing an item that has not been cued (i.e., in no-cue trials, or in trials with a noninformative, neutral cue), thereby producing the so-called *valid retro-cue benefit*. Conversely, testing one of the noncued items in retro-cue trials results in worse performance compared with the no-cue or neutral-cue trials, therefore producing an *invalid retro-cueing cost* (e.g., Astle, Summerfield, Griffin, & Nobre, 2012; Griffin & Nobre, 2003; Pertzov, Bays, Joseph, & Husain, 2013).

Our understanding of the mechanism by which retro-cues improve performance is still very limited. Several tentative explanations have been proposed to account for the retro-cue benefit. In the following section, we describe five mechanisms by which retro-cues could improve performance. Next, we outline the design of the present experiments and how they allow distinguishing between the potential contributions of these mechanisms to the explanation of the retro-cue benefit.

This article was published Online First April 14, 2014.

Alessandra S. Souza, Laura Rerko, and Klaus Oberauer, Department of Psychology, University of Zurich, Zurich, Switzerland.

This research was supported by grants from Forschungskredit of the University of Zurich (project FK-13-083) to the first author, and from the Swiss National Science Foundation (project 100014_126766/1) to the third author. We thank Stefanie Colaço, Iliana Karipidis, Silja Sollberger, and Mirko Thalmann for collecting the data.

Correspondence concerning this article should be addressed to Alessandra S. Souza, Department of Psychology, Cognitive Psychology Unit, University of Zürich, Binzmühlestrasse 14/22, 8050 Zurich, Switzerland. E-mail: a.souza@psychologie.uzh.ch

[1] The causes of the capacity limit are an issue of current debate. Some researchers have argued that it reflects a fixed number of discrete "slots" (Cowan, 2010; Luck & Vogel, 1997); others attribute it to a limited continuously divisible resource (Bays & Husain, 2008; Just & Carpenter, 1992). Our view is that the capacity of WM is limited by interference between items, and, in particular, between bindings of items to their context (Oberauer, 2009; Oberauer, Lewandowsky, Farrell, Jarrold, & Greaves, 2012). The present research was motivated in part by our work on an interference model of WM, but the rationale of our work does not depend on a specific hypothesis about the nature of WM capacity, and therefore we describe it in generic terms.

## Explanations of the Retro-Cue Benefit

At least five potential mechanisms have been raised to account for the performance advantages produced by retro-cues. Table 1 lists these accounts, their basic assumptions, and predictions.

The first explanation states that performance on tests for a retro-cued item is improved because the cued item's representation is enhanced or strengthened (Kuo, Yeh, Chen, & D'Esposito, 2011; Lepsien, Thornton, & Nobre, 2011; Nobre, Griffin, & Rao, 2008; Rerko & Oberauer, 2013). According to this hypothesis, the retro-cued item suffers from less competition from the noncued items in WM, and this leads to faster and more accurate retrieval of the cued item. Strengthening also reduces the set size effect in retro-cue trials, because when the cued item is the strongest among all memory competitors, their number becomes less important. According to this assumption, noncued items are maintained in WM unchanged, but they are less accessible than the cued item (Rerko & Oberauer, 2013). We will refer to this account as the *strengthening hypothesis*.

Alternatively, retro-cues could provide an opportunity to reduce the memory load in retro-cue trials by removing the noncued (currently irrelevant) items from WM. This is the *removal hypothesis* (Oberauer, 2001; Oberauer, Lewandowsky, Farrell, Jarrold, & Greaves, 2012). According to this account, noncued items are actively removed from the central, capacity-limited component of WM,[2] thereby freeing capacity for the maintenance and selection of relevant (cued) information. This process is reflected in the disappearance of the behavioral and neural markers associated with the noncued items. The behavioral marker is the set size effect: Whereas the size of a relevant memory set affects retrieval latencies, the set size effect of an irrelevant set disappears 1 to 2 s after the cue (Oberauer, 2001, 2002, 2005). Further behavioral evidence for removal comes from a recent study by Williams, Hong, Kang, Carlisle, and Woodman (2013), who showed that a retro-cue indicating one of two items in WM as relevant virtually wiped out memory for the remaining item in a surprise test.

Neural markers of WM contents were obtained from pattern classifiers trained to identify the class of stimuli that a person is holding in WM from their brain activity as measured through fMRI or EEG. When people hold stimuli from two different classes in mind, and one class of stimuli is subsequently cued as relevant, the neural activity pattern reflecting the noncued stimulus class drops to baseline (LaRocque, Lewis-Peacock, Drysdale, Oberauer, & Postle, 2013; Lewis-Peacock, Drysdale, Oberauer, & Postle, 2012). Furthermore, in a study by Kuo, Stokes, and Nobre (2012), the contra-lateral delayed activity (CDA; a lateralized marker of WM maintenance) was assessed in neutral-cue and retro-cue trials. In neutral-cue trials and before retro-cueing in retro-cue trials, the CDA was larger when participants maintained four items instead of two items in WM. After retro-cueing, however, the CDA was reduced and converged to a similar level for both set sizes, suggesting that only the cued item was maintained in WM.

Even though removed information no longer affects processing in WM, it is not lost. Noncued information can be retrieved back into WM if it is cued later, as shown by the reappearance of its behavioral and neural signatures (LaRocque et al., 2013; Lewis-Peacock et al., 2012; Oberauer, 2005). This set of findings has motivated the assumption that currently irrelevant items are outsourced from WM, but they can be retrieved back if needed (cf. Oberauer, 2005). These outsourcing and retrieval mechanisms have been recently implemented in a computational model (Oberauer, Souza, Druey, & Gade, 2013).

Although throughout most of this article we will treat strengthening and removal as alternative explanations of the retro-cue benefit, we wish to emphasize that they are not mutually exclusive: The retro-cue benefit could arise from both strengthening of the cued item and removal of the noncued items.

The strengthening hypothesis and the removal hypothesis must be distinguished from three other hypotheses proposed for explaining the retro-cue effect. One of these hypotheses is that the retro-cue protects the cued item from interference by the probe (Makovski, Sussman, & Jiang, 2008). Makovski and colleagues distinguished probe interference from interference between memory items, and argued against the latter because they did not observe a reduction of the set size effect by the retro-cue. Both the strengthening and the removal hypothesis state that the retro-cue reduces interitem interference, by either increasing the strength of the cued item in WM, or by reducing the strength of noncued items, respectively. Therefore, the question whether the retro-cue attenuates the set size effect is important for distinguishing between strengthening and removal, on the one side, and the probe interference hypothesis, on the other side.

The remaining two hypotheses are the *protection* and the *prioritization* hypotheses. The protection account starts from the assumption that, by default, memory representations become degraded during the retention interval, and attention protects the cued item from this degradation (Pertzov et al., 2013). Prioritization means that the cued item receives priority for being compared with the probe at test (Astle et al., 2012; Makovski et al., 2008; Pertzov et al., 2013).

The protection account can explain the retro-cue benefit in comparison with a control condition when the time from encoding to test is kept constant across conditions: Whereas in the control condition memory representations are degraded throughout the retention interval, the retro-cue is assumed to "reduce the rate or probability of degradation" (Matsukura, Luck, & Vecera, 2007, p. 1424; see also Pertzov et al., 2013). The prioritization account can explain the retro-cue benefit in a paradigm requiring comparison of the probe with multiple memory items, such as a change-detection task with a full-array probe display, in which every stimulus in the probe display has to be compared with its counterpart in the remembered memory array, or an item-recognition task, in which a single stimulus must be compared with all memory items (Makovski et al., 2008). Our prioritization account is similar to the one proposed by Matsukura et al. (2007), with the difference that we assume that prioritization can only improve performance when more than one comparison has to be made in memory.

[2] Our theorizing is motivated by a framework that characterizes working memory as consisting of a central, capacity-limited component, and a more encompassing component relying on long-term memory (Cowan, 1999; Oberauer, 2009). For simplicity, we use the term *working memory* to refer only to the capacity-limited central component of the working memory system.

Table 1
*Five Explanations of the Retro-Cue Benefit*

| Hypothesis | Assumptions | Predictions |
|---|---|---|
| Strengthening | The retro-cued item is strengthened compared with the noncued item (which remains unchanged in WM) | (a) Attenuation of the set size effect (E1)<br>(b) Better accessibility of the retro-cued item than an item appended after cueing (E3) |
| Removal | Retro-cued item is maintained in WM, whereas non-cued items are removed from WM, freeing capacity | (a) Attenuation of the set size effect (E1)<br>(c) Reloading benefit (E2; E3) |
| Probe interference | The retro-cued item is protected from interference produced by the presentation of the probe display | (d) No attenuation of the set size effect (E1; E2) |
| Protection | The retro-cued item is protected from degradation (decay; interference) occurring during the retention interval | (e) Retro-cue produces a benefit when time for degradation is longer in no-cue trials compared with retro-cue trials (ruled by design of E1) |
| Prioritization | The retro-cued item is prioritized in the comparison with the probe display | (f) Retro-cue produces a benefit when several comparisons have to be made between memory items and items presented in the probe display (ruled by design in all experiments) |

*Note*. The number of the experiment(s) testing the corresponding prediction is given in parentheses. Some hypotheses were ruled out by the experimental design used, which excluded the contribution of this mechanism to the retro-cue benefit observed. WM = working memory.



## The Present Experiments

The five explanations of the retro-cue benefit make specific predictions regarding the conditions in which the retro-cue benefit should be observed (see Table 1). The present experiments were specifically delineated to adjudicate between differential predictions of these hypotheses (e.g., strengthening, removal, or probe interference) or to rule out some of these explanations (i.e., the protection and prioritization accounts) by experimental design.

According to the strengthening hypothesis, the cued item becomes stronger than noncued items in WM (which are maintained unchanged in WM). The differential strength of the cued item gives it an advantage over its competitors, thereby diminishing the impact of the total number of items in WM. According to the removal hypothesis, noncued items are removed from WM so that they no longer compete with the cued item for WM's limited capacity. Therefore, both the strengthening and the removal hypothesis predict an attenuation of the set size effect by retro-cues. The probe interference hypothesis, in contrast, assumes that the retro-cue serves only to protect the cued item from interference produced by the probe display. According to this account, retro-cues should not attenuate the set size effect.

The evidence speaking to the attenuation of the set size effect by retro-cues is mixed. A number of studies have found that the set size effect observed in the absence of a cue is attenuated when a retro-cue indicates a single item as relevant (Kuo et al., 2012; Nobre et al., 2008) or a subset of items as relevant (LaRocque et al., 2013; Lepsien et al., 2011; Oberauer, 2001, 2002, 2005). Other studies, however, have failed to replicate this reduction of the set size effect by retro-cues (Astle et al., 2012; Makovski et al., 2008).

The studies reporting an attenuation of the set size effect by retro-cues (hereafter *attenuation studies*) and the ones not reporting it (*no-attenuation studies*) differ in at least three aspects. First, attenuation and no-attenuation studies used global and local recognition paradigms, respectively. In global recognition, people are asked whether a single probe stimulus matches any item in the memory set. In this paradigm, the probe stimulus has to be compared with all items in memory in no-cue trials, but only to the cued item in retro-cue trials. In local recognition, people are asked whether a single probe matches a single item in the memory set identified by a contextual cue (e.g., the item's location in an array). In this paradigm, the probe has to be compared only with a single item, regardless of cueing condition, thereby equating the number of comparisons to be made in no-cue and retro-cue trials. This raises the possibility that the attenuation of the set size effect previously reported merely reflects the reduction in the number of comparisons performed in retro-cue trials compared with no-cue trials, as assumed by the prioritization account. Second, attenuation studies used overall longer postcue times between the cue and the probe (>1,000 ms) than some of the no-attenuation studies (400 ms; cf. Makovski et al., 2008). Studies in which one of two subsets of items was cued as relevant have shown that the attenuation of the set size effect develops gradually over time (cf. Oberauer, 2001, 2002). Therefore, it could be that, in some cases, not enough time was provided for people to make full use of the retro-cue to reduce the impact of set size on performance. Third, in the study by Astle et al. (2012), the retro-cues were only 80% valid, which means that in some portion of the trials, a noncued item was tested. This might have reduced the likelihood that people use the retro-cues to fully modulate WM maintenance.

The goal of our first experiment was to examine the conditions in which retro-cues reduce the set size effect. First, we used a local recognition task (aka single-probe change detection) to equate the number of comparisons in no-cue and retro-cue conditions. This feature allowed us to rule out any contribution of prioritization to the retro-cue effect we observed. Second, we used 100% valid retro-cues throughout the experiment, therefore encouraging participants to make strategic use of the retro-cue to improve their WM performance. Third, we varied the postcue time to trace how long it takes to make full use of a retro-cue to modulate WM maintenance. Finally, we compared the retro-cue conditions to a no-cue condition in which the time of testing coincided with the onset of the cue in the retro-cue condition (see Figure 1). Thus, retro-cue and no-cue conditions were equated for the retention interval *before* the cue, and the retro-cue condition had a longer overall retention interval. This design rules out any benefit from protecting the cued item against degradation over time (cf. Makovski et al., 2008, for a similar design). To foreshadow our results,

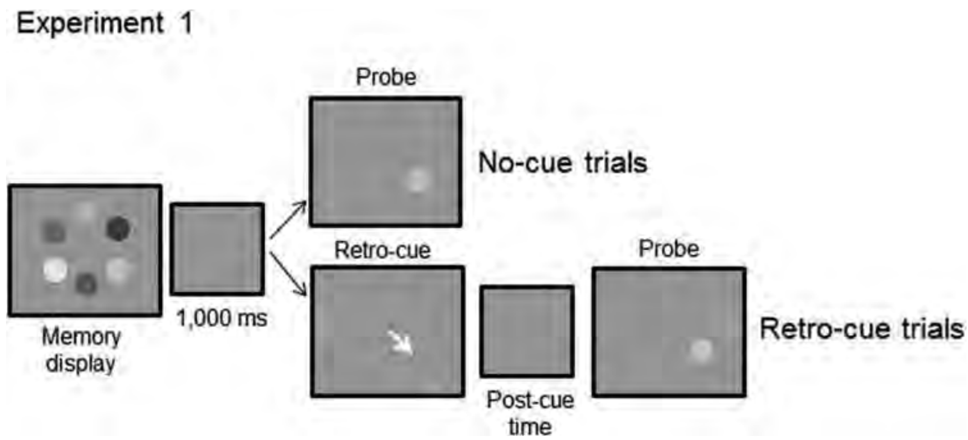


*Figure 1.* Schematic illustration of the no-cue and retro-cue conditions in Experiments 1A and 1B. Participants encoded a memory array with a varied number of items (one, two, four, or six). In the no-cue conditions, a probe stimulus was presented after a 1,000-ms retention interval. In the retro-cue conditions, after the 1,000-ms interval, a retro-cue (a white arrow) was displayed for 100 ms. The retro-cue validly indicated the position of the to-be-probed item. The probe stimulus was presented after a blank postcue interval (whose duration was 100 ms, 400 ms, or 2,000 ms in Experiment 1A; and 100 ms, 400 ms, 1,000 ms, or 2,000 ms in Experiment 1B). The color version of this figure appears in the online article only.



retro-cues attenuated the set size effect both in RTs and accuracy, and this attenuation increased with the postcue time.

Our first experiment rules out probe interference, protection, and prioritization as sufficient explanations of the retro-cue benefit, and supports both the strengthening and the removal hypothesis. The subsequent experiments tested a hypothesis following uniquely from the removal hypothesis, setting it apart from the strengthening hypothesis. According to the removal hypothesis, removing noncued information frees capacity, so that it should be easier to add further material to the capacity-limited part of WM. In contrast, strengthening the cued item (while maintaining the noncued items unchanged) does not free capacity.

## Experiment 1

In this experiment, we tested participants in two experimental tasks in which we varied set size, the presence or absence of a retro-cue, and the time available to make use of the retro-cue before responding to the probe (Experiment 1A = three intervals; Experiment 1B = four intervals).

Our manipulation of the postcue time allowed us to examine whether the amount of time needed to make full use of a retro-cue depends on set size. It is possible that the use of a retro-cue in conditions of high memory load is more demanding, and, therefore, participants take longer to use the retro-cue to improve performance to its maximum. Moreover, if the retro-cue benefit reflects removal of noncued items, it is plausible to assume that removing more items takes longer, so the time needed for achieving the maximum benefit from a retro-cue should increase with set size. Scant research has been devoted to understand the time course of the retro-cue benefit. Tanoue and Berryhill (2012) found that retro-cue benefits emerged for postcue times between 200 and 300 ms, and remained constant thereafter. Set size, however, was not manipulated in their study; hence, it is unknown whether the same amount of time is required to make full use of a retro-cue across different set sizes.

By varying the postcue time, we could trace how retro-cues are used to modulate maintenance of information in WM. One caveat is worth mentioning: Prolonging the postcue time also prolongs the overall retention interval. If this condition is compared with a corresponding no-cue or neutral-cue baseline with an equal overall retention interval, one could not determine whether the retro-cue benefit is related to the time invested in using the retro-cue or, alternatively, to the greater time-based degradation of items in the no-cue trials (cf. Pertzov et al., 2013). To avoid confounding the time for using the retro-cue with the time for memory degradation in no-cue trials, we matched the overall retention interval in the no-cue trials to the precue interval in retro-cue trials (cf. Makovski et al., 2008; see Figure 1). Therefore, our retro-cue trials had overall longer retention intervals compared with the no-cue trials, and any benefits produced by this manipulation could not be attributed to protection of the cued item from memory degradation over time.

### Method

**Participants.** We ran two experiments using a visual recognition task. In Experiment 1A, 20 students of the University of Zurich (16 women; average age = 23 years; range = 20 to 32 years) completed two tasks in counterbalanced order: In one task, participants had to memorize colored disks (data reported in the present article), whereas in the other task, they had to memorize letters (data not reported here). In Experiment 1B, 24 students (20 women; average age = 22.4 years; range = 18 to 28 years) completed a visual recognition task with colored disks. For all experiments reported in this article, participants read and signed an informed consent sheet prior to the study, and were debriefed regarding the purpose of the experiment in the end. Moreover, for all experiments participation was compensated with course credit or monetary reimbursement (15 swiss franks per session).

**Materials and procedure.** All experiments were programmed using MATLAB and the Psychophysics Toolbox exten-

sion (Brainard, 1997; Pelli, 1997). Participants were tested in individual booths, where they sat at approximately 50 cm from the computer screen (viewing distance was unconstrained).

On each trial, a memory display consisting of an array of one, two, four, or six stimuli was presented against a gray background (see Figure 1). The stimuli were colored disks selected from a set of nine distinct colors (i.e., yellow, orange, brown, red, magenta, turquoise, dark blue, light green, and dark green). The memory stimuli for each given set size were arranged evenly spaced around an imaginary circle centered on the middle of the screen. The exact positions of the memory items varied on a trial-by-trial basis: The angle of the first disk was selected randomly (between 1° and 360°) and the remaining items were spaced evenly from that point. Thus, for example, when the set size was four, the interitem distance was 90°, and across trials the exact positions of the items varied (e.g., Trial 1 positions could be 0°, 90°, 180°, and 270°; Trial 2 positions could be 15°, 105°, 195°, and 285°).

Presentation time was set to 150 ms/item, except for the encoding time for Set Size 6 in Experiment 1A, which was 1,000 ms (instead of 900 ms) due to an experimenter error. We presented the items for a rather long time to guarantee that participants had enough time to encode all items to WM. Recall studies have found that with long encoding times, guessing rate is substantially reduced (cf., Bays, Catalao, & Husain, 2009; Bays, Gorgoraptis, Wee, Marshall, & Husain, 2011). Therefore, our choice of long encoding times is conservative in the sense that any performance limitations we observe are likely to be due to limitations in the capacity to maintain or retrieve items from WM.

To discourage verbal coding of the colors, participants were required to perform articulatory suppression (constant repetition of the sequence "*der-die-das*"). Participants were prompted to start articulation in the beginning of each block of trials and to maintain it throughout the block.

As shown in Figure 1, a 1,000-ms blank interval followed the offset of the memory display. In no-cue trials, this interval ended with the presentation of a single probe stimulus in one of the locations (randomly selected) previously occupied by a memory item. In retro-cue trials, this interval ended with the presentation of a white arrow (retro-cue) for 100 ms, which pointed to a randomly selected location in the memory array. The retro-cue validly indicated the position of the probe stimulus. The probe stimulus followed the offset of the retro-cue after a blank postcue time of either 100 ms, 400 ms, or 2,000 ms in Experiment 1A, or 100 ms, 400 ms, 1,000 ms, or 2,000 ms in Experiment 1B. The probe stimulus remained on the screen until participants decided whether it matched the memory item presented in the same location by pressing the left-arrow key for a match response or the right-arrow key for a mismatch response. In 50% of trials, the probe stimulus matched the memory item. In the remaining trials, a mismatch probe was displayed. The mismatch probe was equally likely to be a stimulus not presented in the memory display in the current trial (new probe) or a stimulus presented in a different location of the memory display than the probed one (intrusion probe). We included intrusion probes to force participants to use the information conveyed by the probed position to retrieve the item presented in the same location in the memory array. Visual performance feedback (the word "right" or "wrong" in German) was shown in the center of the screen for 500 ms, and the next trial started 1,500 ms later.

In each experiment, participants completed a total of 1,280 trials across two sessions. In Experiment 1A, there were 80 trials for each Set Size × Postcue Time condition (40 match probes, 20 new probes, and 20 intrusion probes). In Experiment 1B, there were 64 trials for each set size value in each post-cue time condition (32 match probes, 16 new probes, and 16 intrusion probes). In both experiments, the postcue time conditions were presented in a blockwise fashion so that participants could anticipate the postcue time and make optimal use of it. Block order was counterbalanced across participants using a Latin square. Set size and probe type were varied randomly within each block. Prior to the beginning of each block, participants completed 16 practice trials, which were excluded from subsequent analyses. Participants were instructed to respond as fast and as accurately as possible.

## Results

RT and percent correct were the dependent measures of interest. RTs associated with incorrect responses (Experiment 1A = 9.5%; Experiment 1B = 7.3%), faster than 200 ms and slower than 4,000 ms (Experiment 1A = 0.6%; Experiment 1B = 0.2%), as well as those exceeding a participant's mean by more than three standard deviations for each design cell (Experiment 1A = 1.8%; Experiment 1B = 2.0%) were excluded from the RT analyses (altogether, 11.9% and 9.5% of the data points in Experiments 1A and 1B).

Figure 2 shows RTs and percent correct. We included in these analyses responses in all probe-type conditions because the overall pattern of responses across experimental conditions was similar in all probe types. We present RTs and the percentage of correct responses in each probe-type condition in the online supplementary materials.

We were specifically interested in examining whether retrocueing modulates the set size effect, as reflected in the linear relation between set size and performance. Therefore, we contrasted the slopes of the set size functions obtained for each retro-cue condition to the slope of the set size function of the no-cue condition. Given that retro-cues are uninformative when the memory display contains only one item, we removed Set Size 1 from our statistical analyses to avoid the possibility of a misleading Cue Condition × Set Size interaction driven by Set Size 1. Our analytical approach was twofold. First, we ran a repeated measures ANOVA on the data from Experiments 1A and 1B—separately for RT and percent correct—with two independent variables, namely, cue condition and set size. We also ran planned contrasts to follow up on the significant effects found in the ANOVAs. Second, we extracted slope and intercept parameters of the set size effect in each cue condition by fitting a linear mixed effect (LME) model (Pinheiro & Bates, 2000) using the *lme4* function (Bates, Maechler, Bolker, & Walker, 2013) from the R package (R Foundation for Statistical Computing, 2005).

**RT.** The top row of Figure 2 presents the RT in each cue condition as a function of set size separately for Experiment 1A and Experiment 1B. The ANOVA on the RT data of both experiments yielded a significant main effect of cue condition (see Table 2). Simple contrasts (i.e., comparison of each retro-cue condition to the no-cue condition; see Table 3) indicated that responses in all retro-cue conditions were faster than in the no-cue condition, reflecting a retro-cue benefit in RTs.

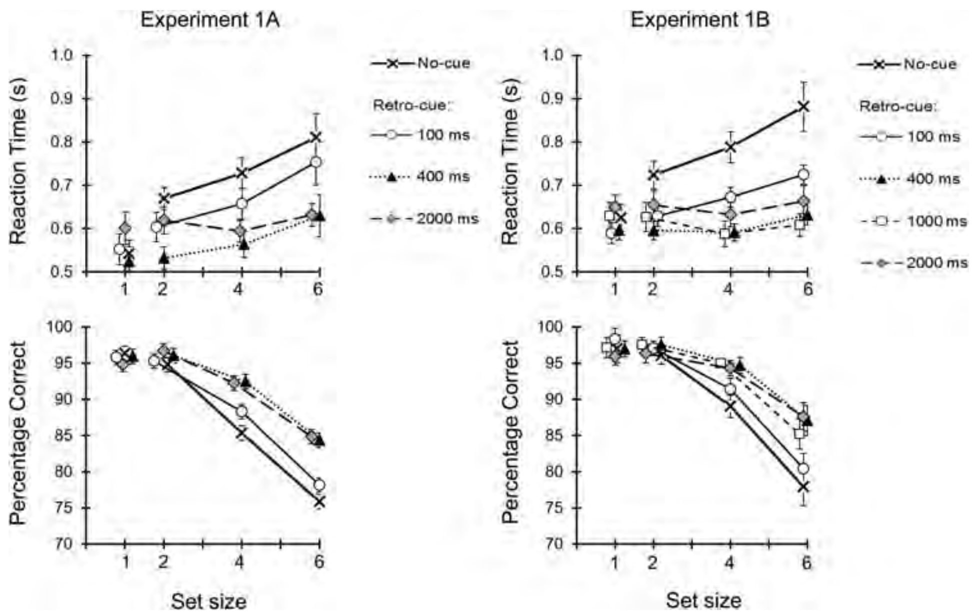


*Figure 2.* Reaction times (top panels) and percent correct (bottom panels) as a function of set size for each experimental condition (no-cue; retro-cue, 100 ms; retro-cue, 400 ms; retro-cue, 1,000 ms; and retro-cue, 2,000 ms) in Experiments 1A and 1B. For Set Size 1, cueing was uninformative (shown in the graphs as unconnected dots). Error bars present 95% within-subjects confidence intervals.

The main effect of set size was significant as well, and linear contrasts showed a significant linear trend indicating slower responses with increases in set size. The set size effect was modulated by an interaction with cue condition, indicating that retro-cues attenuated the effect of set size on RTs. Planned contrasts (see Table 3) showed that this interaction was mostly driven by the retro-cue conditions with longer postcue times (i.e., 2,000 ms) in Experiment 1A; and that this pattern became increasingly stronger with increases in the postcue time in Experiment 1B (with exception of the longest postcue time).

As can be seen in Table 4, the RT set size slopes became increasingly shallow with longer postcue time in the retro-cue trials. To assess whether the slopes obtained for each retro-cue condition significantly differed from each other, we compared nested LME models in which we progressively included fixed effects coding the differences between the cue conditions (see contrast coding in Table 4). We assessed the relative fit of these models by the likelihood ratio test returned by the *lme4* function, the Akaike information criterion, and the Bayesian information criterion (BIC). We report the model fit statistics in Table 5.

As a first step in this analysis, we created a null model for which RT variance was predicted solely by the variables set size (as a fixed effect) and participants (as a random effect). Second, we systematically entered the contrasts in Table 4 as predictors to the model and evaluated whether the resulting extended model improved its fit to the data compared with the model without the added predictor. As shown in Table 4, the first contrast aimed at separating the no-cue from the retro-cue conditions (NC-RC contrast). The second contrast (PCT-linear) orthogonally tested whether the set size effect declined monotonically as a function of

Table 2
*Results of the ANOVAs on the Reaction Time and Percentage Correct Data of Experiments 1A and 1B*

| | Experiment 1A | | | | Experiment 1B | | | |
|---|---|---|---|---|---|---|---|---|
| Variable | *df*s | *F* | *p* | $\eta_p^2$ | *df*s | *F* | *p* | $\eta_p^2$ |
| Reaction time | | | | | | | | |
| Cue condition | (3, 57) | 20.37 | <.001 | .52 | (2.8, 65.3) | 27.99 | <.001 | .55 |
| Set size | (1.1, 20.4) | 21.47 | <.001 | .53 | (1.3, 30.7) | 26.46 | <.001 | .54 |
| Interaction | (3.8, 71.3) | 8.52 | <.001 | .31 | (5.1, 116.3) | 15.34 | <.001 | .40 |
| Percentage correct | | | | | | | | |
| Cue condition | (1.8, 34.6) | 13.8 | <.001 | .42 | (3.2, 72.5) | 25.63 | <.001 | .53 |
| Set size | (1.2, 22.9) | 110.7 | <.001 | .85 | (1.4, 32.5) | 189.14 | <.001 | .89 |
| Interaction | (3.4, 64.8) | 3.1 | .027 | .14 | (4.7, 107.5) | 8.79 | <.001 | .28 |

*Note.* The noninteger degrees of freedom are corrected Greenhouse-Geisser values selected because the sphericity assumption was violated.

Table 3

*Simple Contrasts and Linear Contrasts for the Cue Condition and Set Size Variables, Respectively, on the Reaction Time Data of Experiments 1A and 1B*

| Effect and contrast | Experiment 1A $F(1, 19)$ | $p$ | $\eta_p^2$ | Experiment 1B $F(1, 23)$ | $p$ | $\eta_p^2$ |
|---|---|---|---|---|---|---|
| Cue condition | | | | | | |
| No-cue vs. Retro-cue: 100 ms | 6.75 | .018 | .26 | 40.11 | <.001 | .64 |
| No-cue vs. Retro-cue: 400 ms | 44.93 | <.001 | .70 | 100.10 | <.001 | .81 |
| No-cue vs. Retro-cue: 1,000 ms | | | | 46.32 | <.001 | .67 |
| No-cue vs. Retro-cue: 2,000 ms | 40.34 | <.001 | .68 | 27.35 | <.001 | .54 |
| Set size | | | | | | |
| Linear | 24.01 | <.001 | .56 | 28.16 | <.001 | .55 |
| Quadratic | 7.38 | .014 | .28 | 17.23 | <.001 | .43 |
| Cue Condition × Set Size | | | | | | |
| No-cue vs. Retro-cue: 100 ms | | | | | | |
| Linear | .25 | .623 | .01 | 17.77 | <.001 | .44 |
| Quadratic | .21 | .654 | .01 | .121 | .731 | .01 |
| No-cue vs. Retro-cue: 400 ms | | | | | | |
| Linear | 3.93 | .062 | .172 | 31.56 | <.001 | .58 |
| Quadratic | .08 | .781 | .00 | 2.74 | .111 | .11 |
| No-cue vs. Retro-cue: 1,000 ms | | | | | | |
| Linear | | | | 54.12 | <.001 | .70 |
| Quadratic | | | | 2.82 | .107 | .11 |
| No-cue vs. Retro-cue: 2,000 ms | | | | | | |
| Linear | 31.98 | <.001 | .63 | 39.53 | <.001 | .63 |
| Quadratic | .96 | .339 | .05 | .299 | .590 | .01 |

the postcue time. Given that we had four different postcue times in Experiment 1B, we broke the PCT contrast in two steps for this experiment (PCT-linear and PCT-linear 2).

Table 5 presents the fits for each nested model. For Experiment 1A, the addition of both the NC-RC and the PCT-linear contrasts improved the model's ability to fit the RT data, therefore confirming statistically that the set size slopes were gradually flattened as the postcue time increased. For Experiment 1B, the addition of the NC-RC contrast separating the no-cue from the retro-cue conditions improved model fitting, as well as the addition of the first PCT-linear contrast, which further set apart the *retro-cue:100-ms* condition from the remaining conditions with longer postcue times. The contrast aiming at distinguishing between the three remaining postcue times (PCT-linear 2) did not improve model fitting, suggesting that the slopes of these conditions were not statistically different from each other.

**Percentage correct.** The bottom row of Figure 2 shows the results for the percent-correct measure. We ran the same analyses on the percent-correct data as for the RT. The ANOVA results (see Table 2) show an overall similar pattern to the one obtained for RTs: significant main effects of cue condition and set size, as well as a significant two-way interaction in both experiments.

Simple contrasts (see Table 6) on the cue condition effect indicated that with a very short postcue time (100 ms), overall

Table 4

*Intercept and Slope for the Set Size Curves of Each Cue Condition in Experiments 1A and 1B, and the Dummy Coding Used to Contrast These Conditions*

| | Reaction time | | Percentage correct | | Contrast | | |
|---|---|---|---|---|---|---|---|
| | Intercept | Slope | Intercept | Slope | NC-RC | PCT-linear | PCT-linear 2 |
| Experiment 1A | | | | | | | |
| No-cue | .595 | .035 | 1.044 | −.048 | +1 | 0 | |
| Retro-cue | | | | | | | |
| 100 ms | .522 | .037 | 1.046 | −.043 | −.333 | +.5 | |
| 400 ms | .478 | .024 | 1.029 | −.030 | −.333 | 0 | |
| 2,000 ms | .601 | .003 | 1.033 | −.030 | −.333 | −.5 | |
| Experiment 1B | | | | | | | |
| No-cue | .671 | .042 | 1.058 | −.045 | +1 | 0 | 0 |
| Retro-cue | | | | | | | |
| 100 ms | .607 | .025 | 1.066 | −.042 | −.333 | +1 | 0 |
| 400 ms | .590 | .011 | 1.038 | −.027 | −.333 | −.333 | +.5 |
| 1,000 ms | .643 | −.001 | 1.050 | −.031 | −.333 | −.333 | 0 |
| 2,000 ms | .670 | .005 | 1.016 | −.022 | −.333 | −.333 | −.5 |

*Note.* NC = no-cue condition; RC = retro-cue condition; PCT = postcue time.

Table 5
*Model Fits for the Set Size Slopes From Experiments 1A and 1B*

| Fixed effects | npars | AIC | BIC | LL | Sign. |
|---|---|---|---|---|---|
| Reaction time | | | | | |
| Experiment 1A | | | | | |
| Set Size | 4 | −320.10 | −306.18 | 164.05 | |
| Set Size × NC-RC | 6 | −379.13 | −358.25 | 195.56 | <.001 |
| Set Size × NC-RC + Set Size × PCT-Linear | 8 | −398.79 | −370.95 | 207.40 | <.001 |
| Experiment 1B | | | | | |
| Set Size | 4 | −1092.9 | −1077.4 | 550.47 | |
| Set Size × NC-RC | 6 | −1155.0 | −1131.6 | 583.48 | <.001 |
| Set Size × NC-RC + Set Size × PCT-Linear | 8 | −1202.5 | −1171.4 | 609.24 | <.001 |
| Set Size × NC-RC + Set Size × PCT-Linear + Set Size × PCT-Linear 2 | 10 | −1200.3 | −1161.4 | 610.14 | 0.407 |
| Percentage correct | | | | | |
| Experiment 1A | | | | | |
| Set Size | 4 | −603.65 | −589.73 | 305.82 | |
| Set Size × NC-RC | 6 | −631.93 | −611.04 | 321.96 | <.001 |
| Set Size × NC-RC + Set Size × PCT-Linear | 8 | −647.26 | −619.42 | 331.63 | <.001 |
| Experiment 1B | | | | | |
| Set Size | 4 | −454.35 | −438.80 | 231.17 | |
| Set Size × NC-RC | 6 | −629.65 | −606.34 | 320.83 | <.001 |
| Set Size × NC-RC + Set Size × PCT-Linear | 8 | −651.73 | −620.64 | 333.87 | <.001 |
| Set Size × NC-RC + Set Size × PCT-Linear + Set Size × PCT-Linear 2 | 10 | −649.80 | −610.94 | 334.90 | |

*Note*. Higher log-likelihood values indicate better fits. For AIC and BIC (which are defined as −2*[LL] plus a penalty term for the number of parameters), smaller values indicate better fit of the model. npars = number of free parameters (models with interactions also include both main effects); AIC = Akaike information criterion; BIC = Bayesian information criterion; LL = log likelihood; Sign. = *p* value for significance of the likelihood ratio test comparing each model to the nested model presented in the preceding line.

accuracy improved slightly: The increase in accuracy was marginally significant in Experiment 1A, but significant in Experiment 1B. As the postcue time increased from 100 ms to 2,000 ms, a significant benefit emerged in accuracy, and the *F* values and partial eta-squared values became larger, with exception of the longest postcue time in Experiment 1B (i.e., 2,000 ms). These results suggest that increasing the postcue time increases the likelihood of finding a retro-cue benefit in accuracy.

Regarding set size, the linear contrast showed significant decreases in accuracy as the number of memory items increased. Finally, follow-up contrasts of the Cue Condition × Set Size interaction showed that cueing attenuated the set size effect for

Table 6
*Simple Contrasts and Linear Contrasts for the Cue Condition and Set Size Variables, Respectively, on the Percentage Correct Data of Experiments 1A and 1B*

| | Experiment 1A | | | Experiment 1B | | |
|---|---|---|---|---|---|---|
| Effect and contrasted conditions | $F(1, 19)$ | $p$ | $\eta_p^2$ | $F(1, 23)$ | $p$ | $\eta_p^2$ |
| Cue condition | | | | | | |
| No-cue vs. Retro-cue: 100 ms | 4.02 | .059 | .18 | 6.83 | .016 | .23 |
| No-cue vs. Retro-cue: 400 ms | 13.49 | .002 | .42 | 44.95 | <.001 | .66 |
| No-cue vs. Retro-cue: 1,000 ms | | | | 54.89 | <.001 | .71 |
| No-cue vs. Retro-cue: 2,000 ms | 22.58 | <.001 | .54 | 38.97 | <.001 | .63 |
| Set size | | | | | | |
| Linear | 121.33 | <.001 | .87 | 220.66 | <.001 | .91 |
| Quadratic | 11.71 | <.001 | .38 | 49.85 | <.001 | .68 |
| Cue Condition × Set Size | | | | | | |
| No-cue vs. Retro-cue: 100 ms | | | | | | |
| Linear | .37 | .552 | .02 | .63 | .436 | .03 |
| Quadratic | .38 | .542 | .02 | .166 | .688 | .01 |
| No-cue vs. Retro-cue: 400 ms | | | | | | |
| Linear | 7.35 | .014 | .28 | 17.49 | <.001 | .43 |
| Quadratic | 1.76 | .200 | .09 | .105 | .749 | .01 |
| No-cue vs. Retro-cue: 1,000 ms | | | | | | |
| Linear | | | | 13.53 | .001 | .37 |
| Quadratic | | | | 1.85 | .187 | .08 |
| No-cue vs. Retro-cue: 2,000 ms | | | | | | |
| Linear | 10.59 | .004 | .36 | 40.24 | <.001 | .64 |
| Quadratic | .76 | .396 | .04 | .05 | .819 | .00 |

postcue times longer than 400 ms, and, again, the *F* values and partial eta-squared tended to increase with the postcue time, showing that the attenuation of the set size effect was more robust with long postcue times.

The results shown in Table 4 also indicate that the slopes in the percent-correct measure were flattened by cueing, but to a lesser extent than for the RTs. We went through the same model selection steps as described for the RT analysis. For Experiment 1A, the best-fitting model (see Table 5) for accuracy included both contrasts (NC-RC and PCT-linear) and their interactions with set size, indicating that retro-cues decreased the set size effect on accuracy, and that this effect was modulated by the postcue time. For Experiment 1B, the NC-RC and the PCT-linear contrasts improved the model fitting, but not the addition of the PCT-linear 2 contrast. These results show that retro-cues attenuated the effect of set size in accuracy, and this benefit took at least 400 ms to emerge.

### Discussion

Increasing the set size significantly impaired performance when all items were equally likely to be tested (i.e., in no-cue trials), showing that as load increased, WM capacity was increasingly taxed. When one item of the memory array was retro-cued, the set size effect was reduced, as indicated by the flattening of the set size slopes in RT and percent correct. Our results further qualified the conditions in which attenuation of the set size effect happens: The set size slope was not diminished when the time provided to use the retro-cue was very short (i.e., the 100-ms postcue time), but became increasingly shallow with increasing postcue time. Finally, the attenuation of the set size effect was more evident for RTs than for percent correct.

These findings extend the literature on the retro-cue effect in WM in at least three ways. First, we used a local recognition task requiring the comparison of the probe to only one item in both no-cue and retro-cue trials. This manipulation rules out possible differences in the number of comparisons between no-cue and retro-cue trials as a likely explanation for our attenuation effects. Therefore, we can rule out the contribution of prioritization to our retro-cue benefits.

Second, we showed that the reduction of the set size effect depends on the postcue time: Participants needed time to use the information conveyed by the retro-cue to modulate WM contents (see also Kuo et al., 2011). This helps in explaining the contradictory findings in the literature. As discussed in the introduction, in those studies in which an attenuation of the set size effect by retro-cues was obtained, longer postcue times have been employed (around 1,000 ms). This contrasts with the study of Makovski et al. (2008), in which the postcue time was 400 ms, which might have been too short to allow for the full use of the retro-cue to modulate WM maintenance.

Third, the efficacy of the retro-cue in attenuating the effect of set size was larger for the RT measure, for which the set size slopes were completely flattened, than for the percent correct. The dissociation between these two measures might relate to the different ways in which memory performance is impaired by increasing memory load. On the one hand, the quality of WM representations could be impaired by larger levels of memory load already during encoding and maintenance in WM—stages occurring prior to cueing—which affects the percent-correct measure. Response time, on the other hand, might reflect primarily how representations are selected for retrieval and comparison with the probe—a stage occurring after the retro-cue appears. Arguably, retro-cueing facilitates selection of the cued representation for retrieval, thereby fully counteracting the effect of memory load on speed, but retro-cueing cannot alter the precision with which items have been encoded and maintained in WM before the cue. This interpretation is in line with the results of a recent study by Murray, Nobre, Clark, Cravo, and Stokes (2013), which employed a task assessing the precision of representations in visual WM: Retro-cueing increased the likelihood of retrieving the cued item, but did not alter the precision with which this item was reported in comparison to no-cue trials.

In sum, our results establish that retro-cues can modulate the maintenance of representations in WM. This finding is predicted by both the strengthening and the removal hypothesis, but not by any other hypothesis proposed here. Experiments 2 and 3 contribute to empirically distinguishing between strengthening and removal by testing a unique prediction of the removal hypothesis.

## Experiment 2

The aim of Experiment 2 was to contrast the predictions of the strengthening and removal hypotheses regarding how retro-cues modulate WM maintenance. We used a modified visual recognition paradigm. Participants encoded a memory array (Array 1) comprising either two or four colors and had to respond to a local-recognition probe testing memory for this array (Probe 1). In half of the trials, one item from Array 1 was retro-cued as the one to be probed later, whereas the remaining trials were no-cue trials. In line with the results from Experiment 1, we expect an attenuation of the set size effect when testing memory for Array 1. To test whether this attenuation reflects the increased strength of the cued item relative to its competitors in WM (strengthening hypothesis) or freed capacity (removal hypothesis), we asked participants to encode a second memory array (Array 2, with either two or four colors) during the retention interval of Array 1. Array 2 was presented after the retro-cue in retro-cue trials, but before the probe for Array 1. We tested memory for Array 2 by presenting a probe (Probe 2) subsequently to the response to Probe 1. The sequence of events is illustrated in Figure 3.

Figure 4 illustrates how the design of Experiment 2 allows us to empirically distinguish between strengthening and removal. As shown in the top half of this figure, if the retro-cue is used to strengthen the representation of the cued item compared with the noncued items (which are still drawing on WM's limited capacity), the differential strength of the cued item would make it a strong competitor for retrieval compared with all other items in memory. Therefore, when a second set of items is added to WM and subsequently tested, retrieval of any of these items would suffer from the presence of the strong memory competitor. These predictions would translate in a retro-cue benefit when we test memory for Array 1, but a cost when Array 2 is subsequently tested, because retrieval of the uncued items from Array 2 would suffer increased competition from the now stronger retro-cued item compared with no-cue trials, in which no differential strengthening of any item from Array 1 happened.

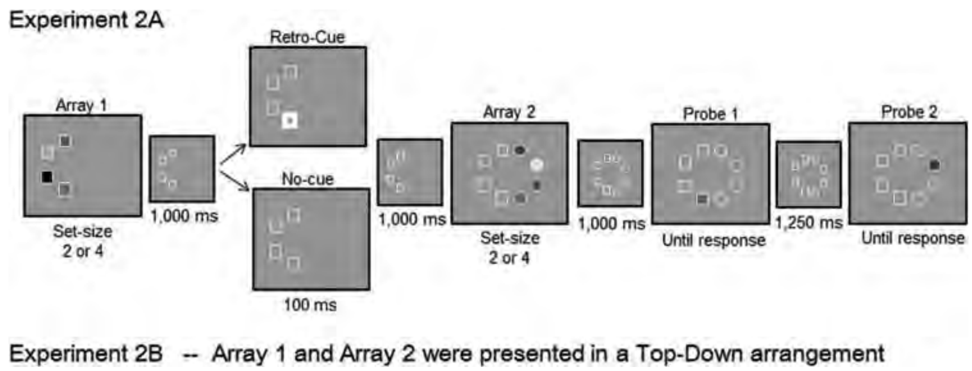


*Figure 3.* Schematic illustration of the trials in Experiment 2A. Participants encoded a memory array (Array 1) comprising either two or four colors presented either in the left or the right hemifield. A second memory array (Array 2; with two or four colors) was displayed in the opposite hemifield 2,100 ms later. During the interval between Array 1 and Array 2, a retro-cue (thicker frame) was displayed in 50% of the trials. The retro-cue validly indicated the location of the item to be probed in Array 1. After presentation of Array 2, a 1,000-ms interval followed. At the end of this interval, an item from Array 1 was probed, followed by a probe from Array 2. Experiment 2B comprised the same conditions. The only difference was that Arrays 1 and 2 were rotated by 90°, resulting in a top-bottom arrangement of the two arrays. The color version of this figure appears in the online article only.

By contrast, the removal hypothesis assumes that the retro-cue indicates which item to maintain and which items to remove from WM. This hypothesis is illustrated in the bottom half of Figure 4, which shows that after a retro-cue, a smaller number of items are maintained in WM, so that that more capacity is available in retro-cue trials. This assumption translates into the prediction that we should observe not only a benefit for the retro-cued item but also a benefit for additional items encoded into WM, such as for a new set appended to WM after retro-cueing.

In a nutshell, Figure 4 shows that the strengthening and the removal hypotheses make different predictions for a set appended to WM after retro-cueing: The strengthening account predicts a cueing cost similar to the invalid retro-cueing costs previously reported in the literature (e.g., Astle et al., 2012; Griffin & Nobre, 2003; Pertzov et al., 2013), whereas the removal hypothesis predicts a cueing benefit. We will refer to this improvement of performance for uncued items appended to WM as a *reloading benefit*.

## Method

**Participants.** Fifty students from the University of Zurich (40 women; average age = 23 years; age range = 18 to 31 years) participated in Experiment 2 (Experiment 2A, $n = 26$; Experiment 2B, $n = 24$). Participants completed one of two experimental versions—Experiment 2A or Experiment 2B—which differed only regarding the arrangement of Array 1 and Array 2 on the display. Three participants (two from Experiment 2A and one from Experiment 2B) were excluded from subsequent analyses due to poor performance (accuracy below 60%, which was three standard deviations below the mean).

**Material and procedure.** We employed a modified color recognition task, which consisted of the sequential presentation of two memory displays and two recognition probes at the end of each trial.

As shown in Figure 3, each trial started with the presentation of a memory display (Array 1) containing two or four colored patches, followed by the presentation of a second memory display (Array 2) also containing either two or four colored patches. The interset interval was 2,100 ms. After the presentation of Array 2, another blank interval of 1,000 ms followed, in the end of which a probe stimulus was shown testing memory for Array 1 (Probe 1), followed by a probe stimulus testing memory for Array 2 (Probe 2). For both arrays, encoding times were 500 ms and 1,000 ms for Set Sizes 2 and 4, respectively. During the time when participants had to maintain Array 1 and Array 2, placeholders (outlines of the memory items) remained on the screen.

There were two types of trials: no-cue trials and retro-cue trials. In no-cue trials, none of the items were cued. In retro-cue trials, one randomly selected item from Array 1 was cued during the interset interval by presenting a thicker frame around its placeholder for 100 ms. The retro-cue was displayed 1,000 ms after the offset of Array 1 and 1,000 ms before the presentation of Array 2. The retro-cue validly indicated the item to be probed from Array 1. Items from Array 2 were never cued.

In the end of the trial, participants had to respond to two local recognition tests: A probe stimulus was shown in one of the locations previously occupied by a memory item. The first presented probe tested memory for Array 1. The probe location was randomly selected in no-cue trials, and was the cued location in retro-cue trials. Participants pressed the left-arrow key or the right-arrow key to indicate a match or mismatch decision, respectively. After responding to Probe 1, visual feedback regarding the correctness of the response was shown for 250 ms, followed by an interprobe interval of 1,000 ms, at the end of which an item from Array 2 was probed (Probe 2). Participants responded to this probe using the same key assignment as for Probe 1. As before, visual performance feedback was provided. The next trial started after 1,000 ms. For both arrays, matching probes were presented in 50% of the trials, and the remaining trials comprised mismatch probes.

Items from Array 1 and Array 2 were distinguished by their shape (square and circle, respectively). The stimuli from both arrays were arranged equally spaced around an imaginary circle

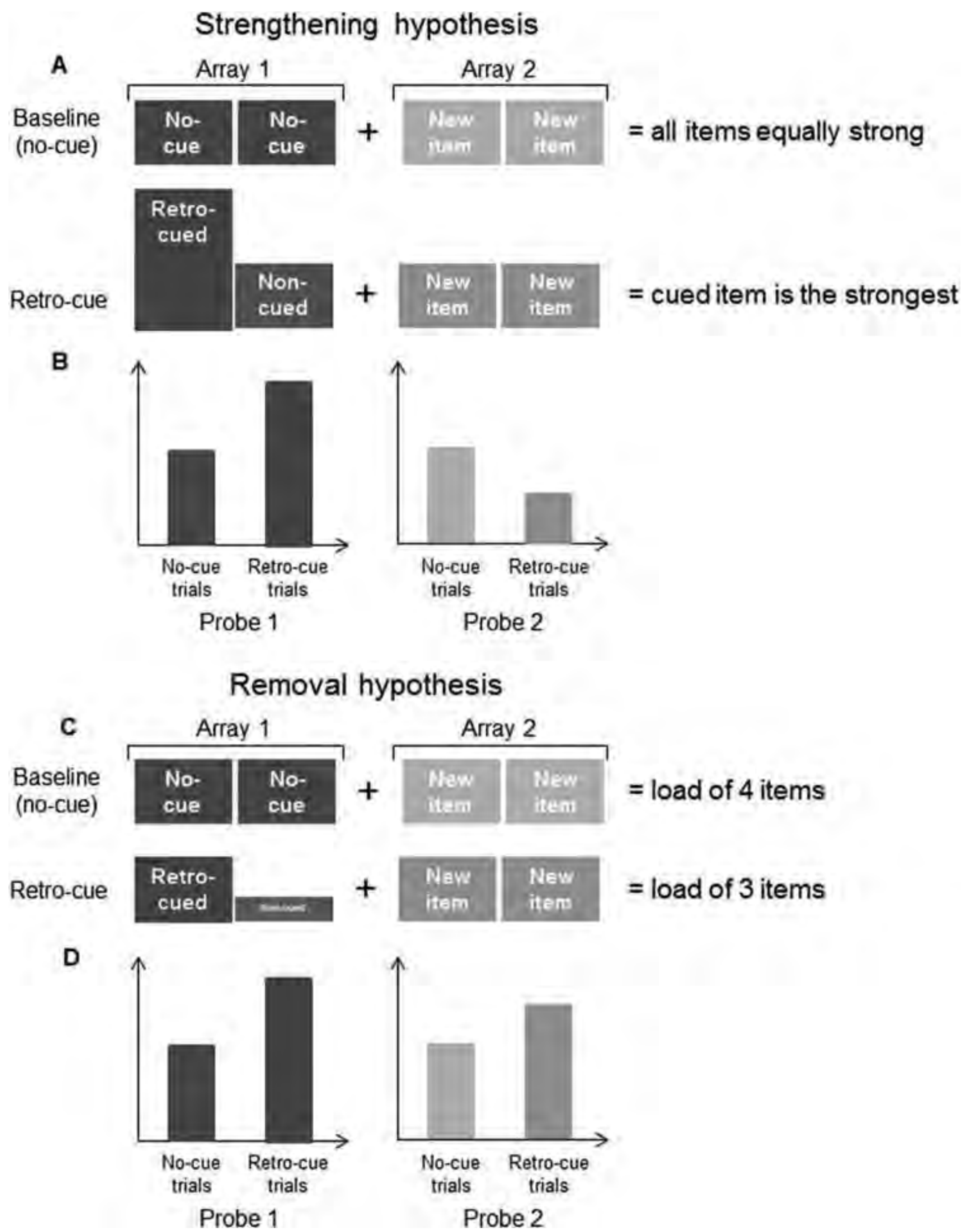


*Figure 4.* Panel A depicts the strengthening hypothesis: The cued item is strengthened without changing the status of the other items in working memory. Panel B illustrates the expected accuracy of responses when testing memory for Array 1 in no-cue and retro-cue trials (Probe 1), and the expected pattern of accuracy in testing memory for Array 2 (Probe 2), according to the assumptions of the strengthening hypothesis: Because the retro-cued item is the strongest, there is a performance benefit in testing this item (dark blue bar), but costs in testing any other item (light blue bar). Panel C depicts the removal hypothesis: Noncued items are removed from working memory, freeing capacity. Panel D shows the expected pattern of responses in testing memory for Array 1 and Array 2: Performance benefits are expected in both memory tests in retro-cue trials compared with no-cue trials because of the lower memory load in these trials. The color version of this figure appears in the online article only.

centered on the middle of the screen. The positions of the memory items were fixed: There were eight selected locations on the screen. Depending on the set size, all of these locations were occupied by memory items, or only a subset of these locations was used. In Experiment 2A, items from Arrays 1 and Array 2 were presented in the left and the right hemifield, respectively (see Figure 3). In Experiment 2B, Arrays 1 and 2 were shifted by 90° degrees, resulting in a top-down arrangement of the displays. Which of the two arrays was displayed in which half of the screen was counterbalanced across participants, but remained constant

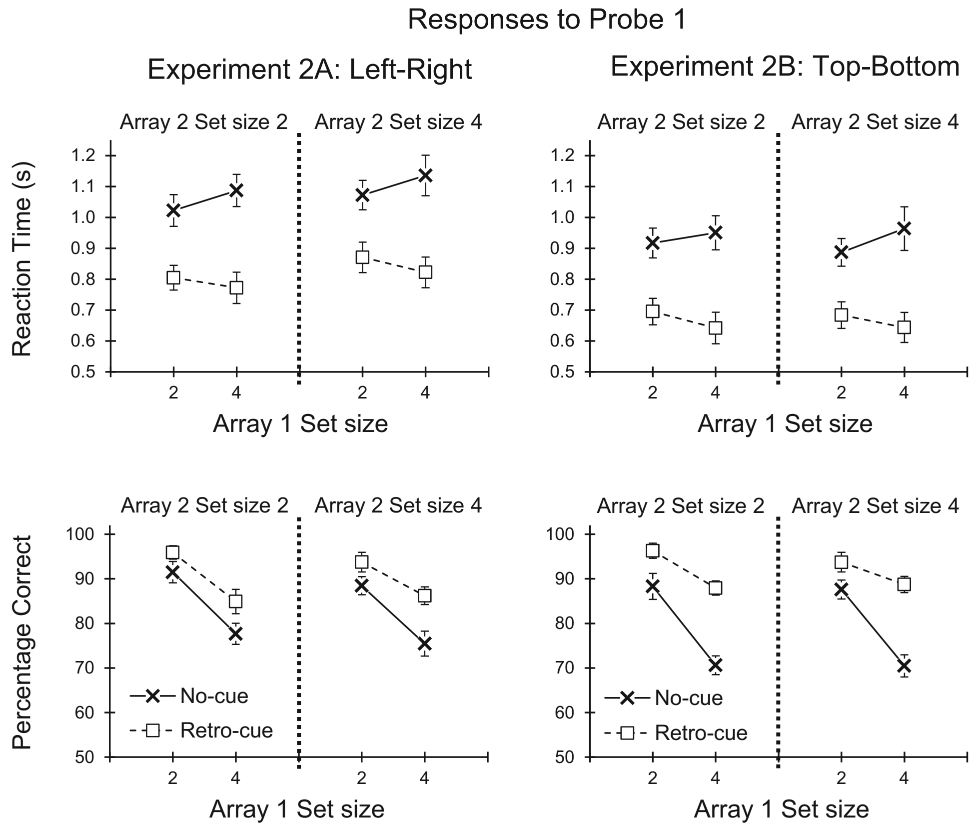


*Figure 5.* Reaction time and percentage of correct responses to Probe 1 in Experiments 2A and 2B. Error bars present 95% within-subjects confidence intervals. Probe 1 always tested an item from Array 1, and therefore we consider the set size of Array 1 to more strongly index the competition in responding to Probe 1 (in our terms, the relevant set size). Participants also had to hold in mind items from Array 2, which were irrelevant to respond to Probe 1 but could not be discarded from memory (currently irrelevant set size). In no-cue trials (cross dots), all items from Array 1 had to be maintained in memory until test, whereas in retro-cue trials (square dot), one item from Array 1 was retro-cued and only this item was relevant to respond to Probe 1.

throughout the experiment for each participant. Colors were selected from a set of 12 distinct colors (i.e., beige, yellow, orange, red, brown, magenta, light green, dark green, turquoise, dark blue, purple, and black). Within the same trial, different colors were selected to be displayed in Arrays 1 and 2, as well as for the two probe displays in case of a new color probe. The set sizes of the two memory arrays were varied randomly and independently across trials, producing four combinations—2_2; 2_4; 4_2; and 4_4—for the set sizes of Array 1 and Array 2, respectively.

Participants completed 640 trials equally divided in two sessions. Each session comprised an equal number of no-cue and retro-cue trials, which were blocked, and the order of the two blocks in one session was counterbalanced across participants. For each cue condition, there were a total of 80 trials with each set size combination (40 match probes, 20 new probes, and 20 intrusion probes). Prior to each block, 12 practice trials were provided. Participants were instructed to respond as fast as possible without sacrificing accuracy.

## Results

RTs associated with incorrect responses (Probe 1: 13.4% and 14.9%; Probe 2: 20.5% and 24.8% in Experiments 2A and 2B, respectively), faster than 200 ms and slower than 6,000 ms (Probe 1: 0.2% and 0.1%; Probe 2: 0.3% and 0.2%), as well as those exceeding a participant's mean by more than three standard deviations in each design cell (Probes 1 and 2: 1.6% in both experiments) were excluded from the RT analyses. Altogether, 15.6% and 16.6% of the responses to Probe 1, and 22.4% and 26.6% of responses to Probe 2, were discarded from the RT analyses for Experiments 2A and 2B, respectively.

In Experiment 2A, responding to Probe 1 was overall slower but more accurate (RT = .943 s; percent correct = 86.6) than responses to Probe 2 (RT = .878 s; percent correct = 79.4), and these differences were statistically significant in paired $t$ tests: RT, $t(23) = 2.16$, $p = .041$; percent correct, $t(23) = 7.75$, $p < .001$. In Experiment 2B, responses to Probe 1 and Probe 2 were about equally fast (RT 1 = .788 s; RT 2 = .775 s), $t(22) = 1.00$, $p = .325$; however, accuracy was again higher for Probe 1 (85.3%) than for Probe 2 (75.2%), $t(22) = 10.45$, $p < .001$.

Figure 5 shows RTs and percent correct to respond to the first presented probe (testing Array 1) in the no-cue and retro-cue conditions as a function of the set size of Array 1. Separate sections within each graph were used to depict the variation in Array 2 set size. Different panels present the results of Experiments 2A and 2B.

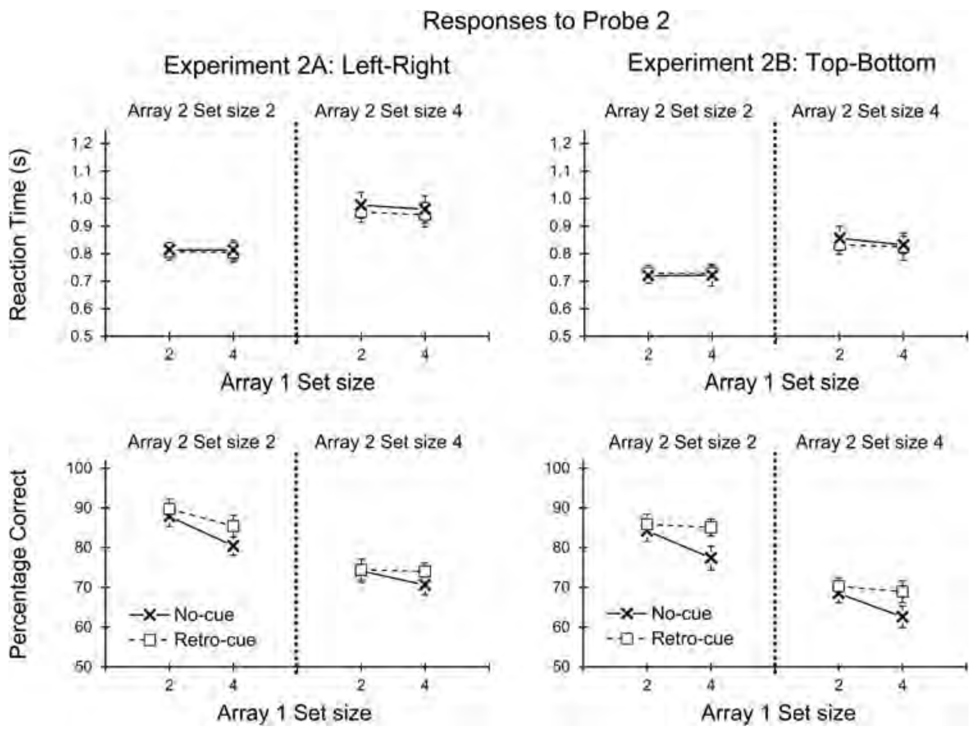


*Figure 6.* Reaction time and percentage of correct responses to Probe 2 in Experiments 2A and 2B. Error bars present 95% within-subjects confidence intervals. Probe 2 always tested an item from Array 2, and therefore we consider the set size of Array 2 to be the relevant set size. At the time of Probe 2, the contents of Array 1 were no longer relevant. None of the items from Array 2 were cued. The division of trials into no-cue and retro-cue refers to the cueing of an item of Array 1.



Figure 6 shows RTs and percent correct to respond to the second presented probe (testing Array 2). Trials were also split into no-cue and retro-cue trials, here referring to the cueing of an item from Array 1 (none of the items from Array 2 were cued). The depiction is similar to the one in Figure 5. As for Experiment 1, in Figures 5 and 6, we included responses for all probe-type conditions because the overall pattern of responses between cueing conditions was similar for all probe types. RTs and the percentage of correct responses to Probe 1 and to Probe 2, respectively, are depicted separately for each probe-type condition in the online supplementary materials.

We were specifically interested in examining whether cueing would attenuate the set size effect for Array 1, and whether cueing an item in Array 1 had an effect on how well Array 2 was encoded and maintained, as reflected in the responses to Probe 2. To this end, we ran ANOVAs separately for our two dependent measures—that is, RT and percent correct—and for the responses to Probe 1 and Probe 2. For each ANOVA (see Table 7), the independent variables were cue condition, Array 1 set size, and Array 2 set size.

**Responses to Probe 1.** These analyses addressed the question whether cueing an item would produce a *retro-cue benefit* and an *attenuation of the set size effect* for Probe 1. Responses to Probe 1 were significantly affected by cueing in both experiments (see Table 7). RTs were 262 ms (Experiment 2A) and 263 ms (Experiment 2B) faster in retro-cue trials compared with non-cue trials. The percentage of correct responses was also higher for retro-cue trials than no-cue trials, with an overall improvement of 6.9% (Experiment 2A) and 12.4% (Experiment 2B). These results indicate the classical retro-cue benefit.

Next, we analyzed the effects of set size and whether cueing attenuated the set size effect. The main effect of the set size of Array 1 was not significant in the RT measure in both experiments, but was highly significant in the percent correct: Performance was reduced by an average of 11.9% (Experiment 2A) and 12.9% (Experiment 2B) when Array 1 set size increased from two to four. Replicating Experiment 1, for both measures and both experiments, cueing significantly interacted with Array 1 set size. In Experiment 2A, cueing benefits in RT increased from 209 ms to 315 ms, and cueing benefits in accuracy increased from 4.8% to 9%, when the set size of Array 1 increased from two to four items. In Experiment 2B, cueing benefits in RT increased from 212 ms to 315 ms when the set size of Array 1 increased from two to four items; and in percent correct, the cueing benefits increased from 6.9% to 17.8%. The increase in the retro-cue benefit with the larger set size is an indication of an attenuation of the set size effect.

In contrast to the main effect of Array 1 set size and the modulation of Array 1 set size by retro-cues, the set size of Array 2 failed to significantly impact responses to Probe 1 (except for the percent correct in Experiment 2A). Furthermore, the interaction between Array 2 set size and cueing, between Array 2 and Array 1 set sizes, and the three-way interaction between all variables yielded nonsignificant results.

Table 7
*Results of the ANOVAs Performed on the Reaction Time and Percentage Correct Data of Experiments 2A and 2B*

| Effect and experiment | *df* | Reaction time *F* | Reaction time *p* | Reaction time $\eta_p^2$ | Percentage correct *F* | Percentage correct *p* | Percentage correct $\eta_p^2$ |
|---|---|---|---|---|---|---|---|
| Responses to Probe 1 | | | | | | | |
| Experiment 2A: Left-Right | | | | | | | |
| Cueing | (1, 23) | 67.92 | <.001 | .78 | 26.71 | <.001 | .54 |
| Array 1 Set Size | (1, 23) | .93 | .345 | .04 | 114.51 | <.001 | .83 |
| Cueing × Array 1 Set Size | (1, 23) | 24.51 | <.001 | .52 | 7.46 | .012 | .25 |
| Array 2 Set Size | (1, 23) | 3.28 | .083 | .13 | 5.19 | .032 | .18 |
| Cueing × Array 2 Set Size | (1, 23) | .18 | .678 | .01 | 3.49 | .075 | .13 |
| Array 1 × Array 2 Set Sizes | (1, 23) | .46 | .595 | .02 | 5.75 | .025 | .20 |
| 3-way interaction | (1, 23) | .27 | .618 | .01 | 1.51 | .232 | .06 |
| Experiment 2B: Top-Bottom | | | | | | | |
| Cueing | (1, 22) | 40.45 | <.001 | .65 | 128.89 | <.001 | .85 |
| Array 1 Set Size | (1, 22) | .10 | .750 | .01 | 184.62 | <.001 | .89 |
| Cueing × Array 1 Set Size | (1, 22) | 16.90 | <.001 | .43 | 40.22 | <.001 | .65 |
| Array 2 Set Size | (1, 22) | .34 | .564 | .02 | 1.03 | .320 | .05 |
| Cueing × Array 2 Set Size | (1, 22) | .19 | .666 | .01 | .21 | .655 | .01 |
| Array 1 × Array 2 Set Sizes | (1, 22) | 1.74 | .200 | .07 | 2.16 | .156 | .09 |
| 3-way interaction | (1, 22) | .40 | .532 | .02 | 1.05 | .317 | .05 |
| Responses to Probe 2 | | | | | | | |
| Experiment 2A: Left-Right | | | | | | | |
| Cueing | (1, 23) | .36 | .555 | .02 | 3.17 | .088 | .12 |
| Array 1 Set Size | (1, 23) | .39 | .538 | .02 | 28.52 | <.001 | .55 |
| Cueing × Array 1 Set Size | (1, 23) | .02 | .895 | .00 | 6.13 | .021 | .21 |
| Array 2 Set Size | (1, 23) | 87.71 | <.001 | .79 | 170.77 | <.001 | .88 |
| Cueing × Array 2 Set Size | (1, 23) | .49 | .491 | .02 | 1.59 | .220 | .07 |
| Array 1 × Array 2 Set Sizes | (1, 23) | .70 | .411 | .03 | 8.72 | .007 | .28 |
| 3-way interaction | (1, 23) | .130 | .721 | .01 | .001 | .979 | .00 |
| Experiment 2B: Top-Bottom | | | | | | | |
| Cueing | (1, 22) | .01 | .930 | .00 | 16.42 | <.001 | .43 |
| Array 1 Set Size | (1, 22) | .69 | .414 | .03 | 16.38 | .001 | .43 |
| Cueing × Array 1 Set Size | (1, 22) | 1.36 | .256 | .06 | 14.71 | .001 | .40 |
| Array 2 Set Size | (1, 22) | 128.54 | <.001 | .85 | 133.70 | <.001 | .86 |
| Cueing × Array 2 Set Size | (1, 22) | .16 | .691 | .01 | .12 | .736 | .01 |
| Array 1 × Array 2 Set Sizes | (1, 22) | 1.41 | .248 | .06 | .00 | .954 | .00 |
| 3-way interaction | (1, 22) | .41 | .528 | .02 | .40 | .532 | .02 |

**Responses to Probe 2.** These analyses test whether there was a reloading benefit for Probe 2. Responses to Probe 2 showed distinct patterns in the RT and percent correct. For RTs, the only factor that significantly impacted responding was the set size of Array 2: Increasing the set size from two to four items slowed responding by 148 ms (Experiment 2A) and 110 ms (Experiment 2B; see Table 7).

Accuracy was significantly affected by the set size of both arrays: Increases in the set size of Array 1 reduced accuracy by 4% (Experiment 2A) and 3.8% (Experiment 2B), whereas increases in the set size of Array 2 reduced accuracy by 12.5% (Experiment 2A) and 15.6% (Experiment 2B).

Cueing an item in Array 1 also improved Array 2 accuracy by 2.8% in Experiment 2A (which yielded a marginally significant effect) and 4.5% in Experiment 2B. This is the predicted reloading benefit. The reloading benefit was modulated by an interaction with the set size of Array 1. In Experiment 2A, responding to Probe 2 was not significantly different across no-cue and retro-cue trials when Array 1 comprised two items, $t(23) = .68$, $p = .501$, but was significantly improved (by 4.2%) when Array 1 contained four items, $t(23) = 2.63$, $p = .015$. The same pattern was observed in Experiment 2B: When Array 1 set size was two, Array 2 accuracy was similar in no-cue and retro-cue trials, $t(22) = 1.72$, $p = .10$, but when Array 1 set size was four, Array 2 accuracy was significantly improved (by 7.1%) by the retro-cue, $t(22) = 4.82$, $p < .001$.

**Relation between the retro-cue benefit and the reloading benefit.** If the attenuation of the set size effect reflects the degree in which WM capacity is freed, then its size should be related to the size of the reloading benefit. This relation is apparent in the interaction between cue condition and Array 1 set size: Retro-cue benefits were larger when Array 1 had four than when it had two items, and, accordingly, the reloading benefits were mainly observed when Array 1 contained four items. To examine this relation more closely, we investigated the correlation between the retro-cue benefit and the reloading benefit across individuals. We computed the retro-cue benefit obtained in testing Probe 1 and the reloading benefit obtained in testing Probe 2 as a function of the Array 1 set size and Array 2 set size in Experiments 2A and 2B. Our goal was to examine whether an individual's size of the retro-cue benefit adds to the prediction of that individual's reloading benefit.

To evaluate these predictions, we built nested LME models, as shown in Table 8. First, we created a null model in which the reloading benefit was estimated only as a function of random effects (i.e., experiment nested with participant). Second, we en-

tered to this model Array 1 set size as a fixed effect. This predictor significantly improved the model fit compared with the null model (see Table 8). Next, we added the size of the retro-cue benefit as a fixed effect. This again improved the model fit. This improvement was significant in the likelihood ratio test, but was only slightly higher in the BIC score, which is a more conservative test. Therefore, the positive correlation between the retro-cue benefit and the reloading benefit is tentatively but not unambiguously supported by our modeling.

### Discussion

In Experiment 2, retro-cuing an item from Array 1 attenuated the set size effect in RTs and percent correct. Replicating the results of Experiment 1, this attenuation was larger for RTs than for percent correct. We speculatively interpret this observation as a dissociation of how set size affects selection of representations for retrieval (as reflected by RTs), and the quality and quantity of representations in memory (as reflected by percent correct). Retro-cues can completely neutralize the competition between memory representations for selection at retrieval. However, their ability to repair the damage inflicted on representations during encoding and maintenance prior to the cue is limited.

In the present set of experiments, we observed that cueing an item from Array 1 produced a reloading benefit when a new set had to be appended to WM. This reloading benefit was related to the amount of items participants could remove from WM in response to cueing: When the cue pointed to one item out of two (therefore allowing removal of only one item), the retro-cue benefit was small, and it did not improve memory for Array 2; however, when the cue pointed to one item out of four (allowing removal of three items), the retro-cue benefit was larger, and memory for Array 2 was improved.

The reloading benefit was limited to the percent-correct measure. The speed with which participants responded to Probe 2 was independent of whether or not there was a retro-cue, and also independent of Array 1 set size. This is not surprising if one considers the interprobe interval used in the current experiments (1,250 ms). After responding to Probe 1, all items from Array 1 became irrelevant and could be removed from WM. Therefore, the interprobe interval could be used to fully remove the now irrelevant Array 1 from WM, completely neutralizing its impact on Probe 2 RTs. Removal of Array 1 after Probe 1 can also explain why RTs to Probe 2 were faster overall than RTs to Probe 1. At the same time, removal of Array 1 after Probe 1 cannot fully counteract the impact of Array 1 on percent correct. Array 1 set size still affected Probe 2 percent correct even after this long interprobe interval, and we could still observe the beneficial effects of retro-cueing on this measure. Taken together, these results confirm our assumption that removing a subset of representations from WM can eliminate those items entirely from the competition at retrieval, thereby neutralizing their effect on RTs, but removal cannot completely undo the damage from memory load on the quality and quantity of representations in WM, which is reflected in accuracy.

One aspect of our procedure in Experiment 2 might seem to limit the conclusion that retro-cues free capacity to encode new items to WM: Responses to Probe 2 were generally less accurate than responses to Probe 1 in no-cue trials. This might suggest that items from Array 2 were not very well represented in WM in the first place. Time-based degradation cannot explain this difference, given that the items from Array 2 were encoded much later during the trial and therefore had to be maintained for a shorter period of time. One likely explanation of this baseline difference relates to output interference. It is known that responding to multiple tests creates output interference reducing performance as items are successively recalled in tests of verbal WM (cf. Farrell & Lewandowsky, 2012; Oberauer, 2003; Tan & Ward, 2007) and visual WM (cf. Woodman & Vecera, 2011). Accordingly, it is possible that responses to Probe 2 were less accurate because of the output interference produced by responding to Probe 1. Because our test of Array 2 was affected by output interference in a way different from our test of Array 1, it is important to ensure that the reloading benefit is not specific to a test of Array 2 subsequent to testing Array 1.

## Experiment 3

In Experiment 3, we used the same paradigm as in Experiment 2 and, in addition, varied the testing order to examine whether the reloading benefit is obtained when Array 2 is tested before testing Array 1. A positive answer would strengthen our conclusion that retro-cues are used to free WM capacity. If the retro-cued item is strengthened relative to the noncued items, as assumed by the strengthening account, then testing an uncued item before testing the cued item should lead to an even stronger chance of finding a cueing cost: The stronger retro-cued item would get in the way of selecting an uncued item. However, if the retro-cue is used to

Table 8
*Statistics for the Nested LME Models Predicting the Reloading Benefit in Experiment 2*

| Fixed effects | npars | AIC | BIC | LL | Sign. |
|---|---|---|---|---|---|
| 0. — | 3 | −386.69 | −376.98 | 196.34 | |
| 1. Array 1 Set Size | 4 | −399.97 | −387.02 | 203.98 | <.001 |
| 2. Array 1 Set Size + RC Benefit | 5 | −404.43 | −388.25 | 207.22 | .011 |
| 3. Array 1 Set Size × RC Benefit | 6 | −404.49 | −385.07 | 208.25 | .151 |

*Note*. Higher log-likelihood values indicate better fits. For AIC and BIC (which are defined as −2*[LL] plus a penalty term for the number of parameters), smaller values indicate better fit of the model. npars = number of free parameters (models with interactions also include both main effects); AIC = Akaike information criterion; BIC = Bayesian information criterion; LL = log likelihood; Sign. = *p* value for significance of the likelihood ratio test comparing each model with the nested model presented in the preceding line; RC = retro-cue.

remove items from WM, and thereby free capacity, then the order of probing should not matter for whether or not we find a reloading benefit.

## Method

**Participants, materials, and procedure.** A new sample of 24 students (12 women; average age = 24.7 years; range = 19 to 33 years) participated in one 1-hr session. The experimental task was the same as the one described for Experiment 2B (i.e., top–bottom presentation of the memory arrays), with two exceptions. First, the set sizes of Array 1 and Array 2 were fixed to four (therefore creating a load of eight items in no-cue trials, and possibly of five items after retro-cueing). Second, we manipulated probe order: In half of the trials, Array 1 was probed before Array 2 (as was done in Experiment 2), whereas in the other half of the trials, Array 2 was probed before Array 1. Therefore, in Experiment 3, there were four conditions created by crossing two variables: cueing (no-cue or retro-cue of an item from Array 1) and probe order (Array 1 probed first and Array 2 probed second, or the reversed order, with Array 2 probed first and Array 1 probed second). There were 80 trials in each experimental condition, which were presented in a blockwise fashion. The order of the conditions was counterbalanced across participants. Prior to each block, participants completed 12 practice trials, which were discarded from subsequent analyses.

## Results

RTs associated with incorrect responses (26.6% of the responses to the probe testing Array 1, and 31% of the responses to the probe testing Array 2) faster than 200 ms and slower than 6,000 ms (0.4% for both types of probes), as well as those exceeding a participant's mean by more than three standard deviations in each design cell (1.5% and 1.3% for probes testing Array 1 and Array 2, respectively), were excluded from the RT analyses. Altogether, 28.5% of the responses to Probe 1 and 32.7% of the responses to Probe 2 were excluded from the RT analyses.

Table 9 presents the mean RT and percent correct. We ran an ANOVA (see Table 10) with the variables cue condition and probe order separately for each dependent measure and for the probes testing Array 1 and Array 2.

When the probe tested memory for Array 1, a significant retro-cue benefit emerged both in RTs (334 ms) and percent correct (14.2%). Probe order did not influence RT or percentage of correct responses to the probe testing Array 1, and there was no interaction between these variables. When the probe tested memory for Array 2, a significant reloading benefit emerged in accuracy (2.8%), but not in RT, replicating the results of Experiment 2. The interaction of probe order and cueing was far from significant, implying that there was no evidence in the data that the reloading benefit was modulated by the order of testing. Contrary to the responses to the probe testing Array 1, probe order had a large main effect on the accuracy of responses to the probe testing Array 2: Responses were, on average, 15% more accurate when Array 2 was tested first than when it was tested second.

## Discussion

In Experiment 3, we obtained a reloading benefit independently of the order in which memory for Array 2 was tested. These results replicate the ones reported in Experiment 2 and support our conclusion that retro-cues free WM capacity. In addition, Experiment 3 showed that retro-cue benefits are obtained regardless of whether Array 1 is tested first or second. This observation replicates a finding by Rerko, Souza, and Oberauer (2014) and confirms that the retro-cue benefit does not require sustained focal attention to the cued item.

The order of probing had an impact on how well participants responded to Array 2: Accuracy was overall higher when Array 2 was tested first than when it was tested second. This finding

Table 9
*Reaction Time and Percentage of Correct Responses to Probes From Array 1 and Array 2 as a Function of the Cue Condition (No-Cue vs. Retro-Cue) and Order in Which the Arrays Were Tested (Probed 1st or Probed 2nd) in Experiment 3*

| | No-cue | | Retro-cue | | |
|---|---|---|---|---|---|
| | Mean | CIs | Mean | CIs | Benefit |
| Array 1 | | | | | |
| Reaction time | | | | | |
| Probed 1st | 1.209 | [1.106–1.312] | 0.849 | [0.776–0.992] | 0.360 |
| Probed 2nd | 1.111 | [1.036–1.186] | 0.790 | [0.703–0.877] | 0.321 |
| Percentage correct | | | | | |
| Probed 1st | 65.4 | [62.1–68.7] | 80.8 | [77.5–84.1] | 15.4 |
| Probed 2nd | 67.3 | [63.4–71.2] | 80.3 | [77.4–83.2] | 13.0 |
| Array 2 | | | | | |
| Reaction time | | | | | |
| Probed 1st | 1.014 | [0.924–1.104] | 1.055 | [0.960–1.150] | −0.041 |
| Probed 2nd | 1.017 | [0.909–1.125] | 1.037 | [0.948–1.126] | −0.020 |
| Percentage correct | | | | | |
| Probed 1st | 75.8 | [72.9–78.7] | 78.1 | [76.0–80.2] | 2.3 |
| Probed 2nd | 59.7 | [56.9–62.5] | 63.0 | [60.5–65.5] | 3.3 |

*Note*. CIs = 95% within-subjects confidence intervals.

Table 10
*Results of the ANOVAs Performed on the Reaction Time and Percentage of Correct Responses to the Probe Testing Memory for Array 1 and the Probe Testing Memory for Array 2 in Experiment 3*

| | Reaction time | | | Percentage correct | | |
|---|---|---|---|---|---|---|
| Variable | $F(1, 23)$ | $p$ | $\eta_p^2$ | $F(1, 23)$ | $p$ | $\eta_p^2$ |
| Array 1 | | | | | | |
| Cueing | 79.31 | <.001 | .78 | 46.66 | <.001 | .67 |
| Probe order | 1.65 | .212 | .07 | .197 | .662 | .01 |
| Cueing × Probe Order | .571 | .458 | .02 | .868 | .361 | .04 |
| Array 2 | | | | | | |
| Cueing | .700 | .411 | .03 | 5.54 | .028 | .19 |
| Probe order | .012 | .914 | .00 | 98.91 | <.001 | .81 |
| Cueing × Probe Order | .206 | .654 | .00 | .210 | .651 | .01 |

confirms the role of output interference in visual WM. Output interference affected only Array 2. This pattern matches previous observations of an interaction between input position and output position (Cowan, Saults, Elliott, & Moreno, 2002; Jones & Oberauer, 2013). The reason for this interaction is, thus far, not well understood; fortunately, it is not important for the present purpose. Experiment 3 deconfounded the effect of output order from the effects of interest—the retro-cue benefit and the reloading benefit—and replicated both these effects.

Experiment 3 also offers the opportunity to compare performance between Array 1 and Array 2 because this comparison is no longer confounded with probe order. An account of the retro-cue benefit based exclusively on removal of the noncued items implies that the strength of the representation of the cued item in Array 1 following a retro-cue is the same as the strength of all Array 1 representations in the absence of a cue. Moreover, mean performance (averaged across probe orders) on tests of Array 1 and of Array 2 in the no-cue condition was approximately the same in this experiment (Array 1 = 66.1%; Array 2 = 67.8%), implying that memory items were initially encoded with about equal strength for both arrays. A pure removal account therefore implies that a cued item in Array 1 is represented with the same strength as each of the uncued items in Array 2. On that basis, we should expect that in the retro-cue condition accuracy for the test of the cued item in Array 1 should be the same as accuracy in testing a randomly selected item from Array 2. Contrary to that expectation, accuracy on the cued item in Array 1 (80.6%) far exceeds accuracy in Array 2 (70.6%) in the retro-cue condition. This difference implies that the retro-cue benefit does not just arise because of the lower memory load on retro-cue trials, because the memory load is the same for the cued item in Array 1 and the Array 2 items. The improved performance for the cued Array 1 item could reflect a contribution of strengthening to the retro-cue benefit. Alternatively, it could arise because the cued item in Array 1 remains spatially isolated, because the items spatially close to it were removed (in line with the removal account). The latter would be in accordance with studies showing that people's tendency to confuse items with other items from the array increases with their spatial proximity (e.g., Rerko, Oberauer, & Lin, 2014).

## General Discussion

Across three sets of experiments, we consistently found evidence that retro-cues counteract the effects of memory load, resulting in an attenuation of the set size effect on RTs and accuracy. Furthermore, this attenuation of set size effect by retro-cues reflects a relatively slow process: In order to fully use a retro-cue to modulate WM maintenance, participants needed more than 400 ms but not more than 1 s. Our results are in agreement with the assumption that retro-cueing allows a gradual process to take place by which some contents (cued, relevant ones) are foregrounded over others (noncued, currently irrelevant ones).

We went one step further in trying to understand the possible mechanisms by which retro-cues are used to modulate WM maintenance. We tested two mechanisms by which the effect of memory load could be reduced for the cued information: strengthening of the cued item or removal of noncued items. According to the strengthening hypothesis, an item in WM is strengthened when a retro-cue directs attention to it. According to the removal hypothesis, the retro-cue allows for the removal of irrelevant information from the central, capacity-limited part of WM, thereby freeing WM capacity (cf. Oberauer, 2001, 2002; Oberauer et al., 2012). Crucially, previous studies have shown that the noncued information is not lost but can be brought back to WM if cued later, as shown by the reappearance of its behavioral and neural signatures (LaRocque et al., 2013; Lewis-Peacock et al., 2012; Oberauer, 2005). Arguably, removing the active traces of the noncued information frees WM capacity, enabling more accurate maintenance and more efficient retrieval of the currently relevant information, and encoding of new potentially relevant information.

In Experiments 2 and 3, we tested a unique prediction from the removal hypothesis: If noncued items are removed from WM, this should free WM capacity. Our results were in agreement with this prediction: We found a reloading benefit when participants could get rid of a large amount of items in WM.

Can the removal hypothesis be reconciled with the findings of previous studies that have been interpreted as challenging for the assumption of removal? For example, in a recent study by Rerko and Oberauer (2013, Experiment 1), trials without a cue, trials with a single retro-cue, and trials with two subsequent retro-cues were randomly intermixed, and the last presented retro-cue was valid in predicting the item to be probed later. Because participants could not predict how many cues they will see on a given trial, Rerko and Oberauer argued that they could not remove noncued items after the first cue. This unpredictable condition was compared with a predictable condition in which only no-cue and single-cue trials were presented, such that participants could rely 100% on the first presented retro-cue. Thus, they could remove the noncued items upon seeing the first cue. Nevertheless, single-cue trials in the unpredictable and the predictable condition produced equal retro-cue benefits. In another experiment, multiple cues were presented in rapid succession (up to three cues), and the last cue was the one validly indicating the item to compare to the probe. In this experiment, single-, two-, and three-cue trials produced comparable performance, and cueing a previously cued item again (A-B-A sequence) yielded better performance than cueing three

different items in succession (C-B-A sequence). Rerko and Oberauer (2013) interpreted these findings as evidence that (a) focal attention can be switched flexibly among WM items; (b) focusing attention on the cued item strengthens that item in WM without changing the state of the noncued items, which remain available to be focused later; and (c) a previously cued, but now irrelevant, item remains strengthened, such that when a cue points to it again (as in the A-B-A sequence), it can build on its residual activation. This interpretation appears to clash with the idea that noncued items are removed from WM.

We believe, however, that these results can be reconciled with the removal hypothesis as follows. First, it is possible that participants started removing noncued items upon presentation of the first cue in all trials because noncued items could always be brought back to WM if cued later, as shown in several experiments using behavioral and neurophysiological measures (LaRocque et al., 2013; Lewis-Peacock et al., 2012; Oberauer, 2005). Therefore, the lack of a difference between the predictable and the unpredictable conditions in Rerko and Oberauer (2013) might simply reflect the fact that removal takes place in both conditions. Second, the present experiments, in agreement with previous studies (Oberauer, 2001, 2002), show that removal is a gradual process. The 500- to 700-ms intercue intervals in the experiments of Rerko and Oberauer (2013) might therefore not have been sufficient for complete removal. When a first cue is followed by a second cue within 500 to 700 ms, the item cued second is not yet completely removed from WM and therefore could be restored quickly. This could explain the comparable levels of performance for single-cue, two-cue, and even three-cue trials. Moreover, when a previously cued item is later cued again (as in the A-B-A sequence), it would have been removed to a lesser degree than an item not cued previously (as in the C-B-A sequence). Therefore, the third cue in the A-B-A sequence can build on more residual memory strength of item A than in the C-B-A sequence.

Finally, strengthening and removal are not necessarily incompatible processes. It is possible that the cued item is strengthened and noncued items are removed, and the joint effect of both processes gives rise to the retro-cue benefit and the attenuation of the effect of memory load. The present experiments strongly imply that removal of noncued items plays a role in the retro-cue effect, but they, by no means, rule out an additional role of strengthening of cued items. The finding in Experiment 3 that accuracy on tests for the cued item exceeds what would be expected on a pure removal account may suggest that strengthening of the cued item contributes to the retro-cue effect. If this is the case, though, the contribution of strengthening is arguably smaller than the contribution of removal: As illustrated in Figure 4, strengthening by itself predicts a cueing cost for Array 2, whereas removal predicts a benefit. The fact that we observed a net benefit—the reloading benefit—implies that removal of noncued items must have weighed more strongly than strengthening of the cued item.

The reduction of memory load we observed could be explained either by an active removal process (Oberauer et al., 2012) or, alternatively, by a passive process in which noncued representations decay during the postcue interval and eventually drop out of WM, thereby reducing the degree to which they compete with the cued item and freeing up capacity for encoding further items. Whereas the active removal hypothesis unambiguously predicts the reloading benefit we observed, it is not clear that this passive decay account predicts it. According to the passive decay hypothesis, the retro-cued item is maintained in WM through protection, whereas the noncued items drop out, resulting in a load of one item plus an increasingly diminishing residual of the noncued items. In the no-cue condition, one of three things could happen: (a) all items drop out, resulting in a memory load less than in the retro-cue condition; (b) all items drop out except for one randomly selected item that is protected by attention, resulting in a load equal to the retro-cue condition; or (c) more than one item is protected, resulting in a load higher than in the retro-cue condition. Only the third scenario would imply the prediction of a reloading benefit. Thus, whereas this passive decay account is compatible with the reloading benefit when combined with appropriate additional assumptions, the active removal account directly predicts the reloading benefit.

One way to adjudicate between these two subtly different hypotheses is to investigate whether memory representations do become weaker over time in situations that do not motivate active removal. This question has been debated for several decades in the literature on short-term memory and WM, and the balance of evidence is by now strongly against the notion of gradual decay or deterioration of WM representations over time (Lewandowsky, Oberauer, & Brown, 2009; Oberauer & Lewandowsky, 2013; Shipstead & Engle, 2013; Souza & Oberauer, in press). Therefore, the assumption of active removal of no longer relevant representations from WM has more credibility than the assumption of passive degradation or decay.

## Conclusion

Our results support the conclusion that the contents of WM can be effectively modulated by retro-cues: Cued items are maintained and probably strengthened in WM, whereas noncued items are gradually removed from WM, freeing otherwise used capacity. Across three sets of experiments, removal of noncued items was evidenced by (a) the attenuation of the set size effect, and (b) better memory for a new set of items added to WM after a retro-cue.

## References

Astle, D. E., Summerfield, J., Griffin, I., & Nobre, A. C. (2012). Orienting attention to locations in mental representations. *Attention, Perception, & Psychophysics, 74,* 146–162. doi:10.3758/s13414-011-0218-3

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2013). *lme4: Linear mixed-effects models using Eigen and S4. R package version 1.0–5.* Retrieved from http://cran.r-project.org/web/packages/lme4/index.html

Bays, P. M., Catalao, R. F. G., & Husain, M. (2009). The precision of visual working memory is set by allocation of a shared resource. *Journal of Vision, 9,* 1–11. doi:10.1167/9.10.7

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,* 1–15. doi:10.1167/11.10.6

Bays, P. M., & Husain, M. (2008). Dynamic shifts of limited working memory resources in human vision. *Science, 321,* 851–854. doi:10.1126/science.1158023

Brainard, D. H. (1997). The psychophysics toolbox. *Spatial Vision, 10,* 433–436. doi:10.1163/156856897X00357

Cowan, N. (1999). An embedded-process model of working memory. In A. Miyake & P. Shah (Eds.), *Models of working memory: Mechanisms of active maintenance and executive control* (pp. 62–101). New York, NY: Cambridge University Press. doi:10.1017/CBO9781139174909.006

Cowan, N. (2010). The magical mystery four: How is working memory capacity limited, and why? *Current Directions in Psychological Science, 19,* 51–57. doi:10.1177/0963721409359277

Cowan, N., Saults, J. S., Elliott, E. M., & Moreno, M. V. (2002). Deconfounding serial recall. *Journal of Memory and Language, 46,* 153–177. doi:10.1006/jmla.2001.2805

Farrell, S., & Lewandowsky, S. (2012). Response suppression contributes to recency in serial recall. *Memory & Cognition, 40,* 1070–1080. doi:10.3758/s13421-012-0212-6

Gazzaley, A., & Nobre, A. C. (2012). Top-down modulation: Bridging selective attention and working memory. *Trends in Cognitive Sciences, 16,* 129–135. doi:10.1016/j.tics.2011.11.014

Griffin, I. C., & Nobre, A. C. (2003). Orienting attention to locations in internal representations. *Journal of Cognitive Neuroscience, 15,* 1176–1194. doi:10.1162/089892903322598139

Jones, T., & Oberauer, K. (2013). Serial-position effects for items and relations in short-term memory. *Memory, 21,* 347–365. doi:10.1080/09658211.2012.726629

Just, M. A., & Carpenter, P. A. (1992). A capacity theory of comprehension: Individual differences in working memory. *Psychological Review, 99,* 122–149. doi:10.1037/0033-295X.99.1.122

Kuo, B.-C., Stokes, M. G., & Nobre, A. C. (2012). Attention modulates maintenance of representations in visual short-term memory. *Journal of Cognitive Neuroscience, 24,* 51–60. doi:10.1162/jocn_a_00087

Kuo, B.-C., Yeh, Y.-Y., Chen, A. J.-W., & D'Esposito, M. (2011). Functional connectivity during top-down modulation of visual short-term memory representations. *Neuropsychologia, 49,* 1589–1596. doi:10.1016/j.neuropsychologia.2010.12.043

Landman, R., Spekreijse, H., & Lamme, V. A. F. (2003). Large capacity storage of integrated objects before change blindness. *Vision Research, 43,* 149–164. doi:10.1016/S0042-6989(02)00402-9

LaRocque, J. J., Lewis-Peacock, J. A., Drysdale, A. T., Oberauer, K., & Postle, B. R. (2013). Decoding attended information in short-term memory: An EEG study. *Journal of Cognitive Neuroscience, 25,* 127–142. doi:10.1162/jocn_a_00305

Lepsien, J., Thornton, I., & Nobre, A. (2011). Modulation of working-memory maintenance by directed attention. *Neuropsychologia, 49,* 1569–1577. doi:10.1016/j.neuropsychologia.2011.03.011

Lewandowsky, S., Oberauer, K., & Brown, G. D. A. (2009). No temporal decay in verbal short-term memory. *Trends in Cognitive Sciences, 13,* 120–126. doi:10.1016/j.tics.2008.12.003

Lewis-Peacock, J. A., Drysdale, A. T., Oberauer, K., & Postle, B. R. (2012). Neural evidence for a distinction between short-term memory and the focus of attention. *Journal of Cognitive Neuroscience, 24,* 61–79. doi:10.1162/jocn_a_00140

Luck, S. J., & Vogel, E. K. (1997). The capacity of visual working memory for features and conjunctions. *Nature, 390,* 279–281. doi:10.1038/36846

Makovski, T., Sussman, R., & Jiang, Y. V. (2008). Orienting attention in visual working memory reduces interference from memory probes. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 34,* 369–380. doi:10.1037/0278-7393.34.2.369

Matsukura, M., Luck, S. J., & Vecera, S. P. (2007). Attention effects during visual short-term memory maintenance: Protection or prioritization? *Perception & Psychophysics, 69,* 1422–1434. doi:10.3758/BF03192957

Murray, A. M., Nobre, A. C., Clark, I. A., Cravo, A. M., & Stokes, M. G. (2013). Attention restores discrete items to visual short-term memory. *Psychological Science, 24,* 550–556. doi:10.1177/0956797612457782

Nobre, A. C., Griffin, I. C., & Rao, A. (2008). Spatial attention can bias search in visual short-term memory. *Frontiers in Human Neuroscience, 1,* 4. doi:10.3389/neuro.09.004.2007

Oberauer, K. (2001). Removing irrelevant information from working memory: A cognitive aging study with the modified Sternberg task. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 27,* 948–957. doi:10.1037/0278-7393.27.4.948

Oberauer, K. (2002). Access to information in working memory: Exploring the focus of attention. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 28,* 411–421. doi:10.1037/0278-7393.28.3.411

Oberauer, K. (2003). Understanding serial position curves in short-term recognition and recall. *Journal of Memory and Language, 49,* 469–483. doi:10.1016/S0749-596X(03)00080-9

Oberauer, K. (2005). Control of the contents of working memory: A comparison of two paradigms and two age groups. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 31,* 714–728. doi:10.1037/0278-7393.31.4.714

Oberauer, K. (2009). Design for a working memory. *The Psychology of Learning and Motivation, 51,* 45–100. doi:10.1016/S0079-7421(09)51002-X

Oberauer, K., & Lewandowsky, S. (2013). Evidence against decay in verbal working memory. *Journal of Experimental Psychology: General, 142,* 380–411. doi:10.1037/a0029588

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. doi:10.3758/s13423-012-0272-4

Oberauer, K., Souza, A. S., Druey, M. D., & Gade, M. (2013). Analogous mechanisms of selection and updating in declarative and procedural working memory: Experiments and a computational model. *Cognitive Psychology, 66,* 157–211. doi:10.1016/j.cogpsych.2012.11.001

Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: Transforming numbers into movies. *Spatial Vision, 10,* 437–442. doi:10.1163/156856897X00366

Pertzov, Y., Bays, P. M., Joseph, S., & Husain, M. (2013). Rapid forgetting prevented by retrospective attention cues. *Journal of Experimental Psychology: Human Perception and Performance, 39,* 1224–1231. doi:10.1037/a0030947

Pinheiro, J. C., & Bates, D. (2000). *Mixed effects models in S and S-Plus*. New York, NY: Springer. doi:10.1007/978-1-4419-0318-1

Rerko, L., & Oberauer, K. (2013). Focused, unfocused, and defocused information in working memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 39,* 1075–1096. doi:10.1037/a0031172

Rerko, L., Oberauer, K., & Lin, H.-Y. (2014). Spatial transposition gradients in visual working memory. *Quarterly Journal of Experimental Psychology (2006), 67,* 3–15. doi:10.1080/17470218.2013.789543

Rerko, L., Souza, A. S., & Oberauer, K. (2014). Retro-cue benefits in working memory without sustained focal attention. *Memory & Cognition*. Advance online publication. doi:10.3758/s13421-013-0392-8

R Foundation for Statistical Computing. (2005). *R: A language and environment for statistical computing (version 2.2.1.)*. Vienna, Austria: Author.

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. doi:10.1037/a0028467

Souza, A. S., & Oberauer, K. (in press). Time-based forgetting in visual working memory reflects temporal distinctiveness, not decay. *Psychonomic Bulletin & Review*.

Sternberg, S. (1966). High-speed scanning in human memory. *Science, 153,* 652–654. doi:10.1126/science.153.3736.652

Sternberg, S. (1975). Memory scanning: New findings and current controversies. *Quarterly Journal of Experimental Psychology, 27,* 1–32. doi:10.1080/14640747508400459

Tan, L., & Ward, G. (2007). Output order in immediate serial recall. *Memory & Cognition, 35,* 1093–1106. doi:10.3758/BF03193481

Tanoue, R. T., & Berryhill, M. (2012). The mental wormhole: Internal attention shifts without regard for distance. *Attention, Perception, & Psychophysics, 74,* 1199–1215. doi:10.3758/s13414-012-0305-0

Williams, M., Hong, S. W., Kang, M.-S., Carlisle, N. B., & Woodman, G. F. (2013). The benefit of forgetting. *Psychonomic Bulletin & Review, 20,* 348–355. doi:10.3758/s13423-012-0354-3

Woodman, G. F., & Vecera, S. P. (2011). The cost of accessing an object's feature stored in visual working memory. *Visual Cognition, 19,* 1–12. doi:10.1080/13506285.2010.521140

Received May 7, 2013
Revision received February 14, 2014
Accepted February 18, 2014 ■