Ann. N.Y. Acad. Sci. ISSN 0077-89

ANNALS OF THE NEW YORK ACADEMY OF SCIENCES
Special Issue: *Attention in Working Memory*
ORIGINAL ARTICLE

Behavioral decoding of working memory items inside and outside the focus of attention

Remington Mallett and Jarrod A. Lewis-Peacock

Department of Psychology, University of Texas at Austin, Austin, Texas

Address for correspondence: Remington Mallett, Department of Psychology, University of Texas at Austin, Seay 2.218, Austin, TX 78712. remym@utexas.edu

How we attend to our thoughts affects how we attend to our environment. Holding information in working memory can automatically bias visual attention toward matching information. By observing attentional biases on reaction times to visual search during a memory delay, it is possible to reconstruct the source of that bias using machine learning techniques and thereby behaviorally decode the content of working memory. Can this be done when more than one item is held in working memory? There is some evidence that multiple items can simultaneously bias attention, but the effects have been inconsistent. One explanation may be that items are stored in different states depending on the current task demands. Recent models propose functionally distinct states of representation for items inside versus outside the focus of attention. Here, we use behavioral decoding to evaluate whether multiple memory items—including temporarily irrelevant items outside the focus of attention—exert biases on visual attention. Only the single item in the focus of attention was decodable. The other item showed a brief attentional bias that dissipated until it returned to the focus of attention. These results support the idea of dynamic, flexible states of working memory across time and priority.

Keywords: attention; working memory; attentional capture; visual search; pattern classification

Introduction

What is on our mind influences what we attend to in our environment. Evidence for this comes from findings in which external attention is automatically biased toward information matching the contents of working memory. People are slower to respond on a visual search task when an item matching an item held in working memory is presented as a distractor[1] (attentional capture). A broad range of mental states and processes can bias our attention, including visual working memory,[2,3] verbal working memory,[4–7] long-term memory,[8–11] associative knowledge,[12,13] implicit memory,[14] and reward.[15] Studies have found converging evidence for attentional capture when the working memory load is held to a single item (although see Refs. 16 and 17), which has spurred important discussions about the overlap of working memory and attention.[18–21]

Previous models of visual working memory[22] have been built around experiments suggesting that only a single representation is capable of biasing attention.[23–26] These findings have recently been challenged by studies reporting attentional capture from multiple working memory items.[27–31] However, there is no clear consensus on this issue. Notably, these studies vary in their experimental procedures and definitions of what constitutes "multiple items" in working memory. Some use a single memory item and a search target that varies by trial[23,24] and is therefore considered to be in working memory.[32,33] Others use the same logic with equal-priority search targets,[27,28] while still others use two working memory items of equal priority.[26,29,30]

While much theoretical debate is devoted to working memory capacity limitations,[34–36] there is also emphasis on dynamic representational states within working memory that are differentiated not only by capacity limits but also by priority and goal relevance.[37–41] Indeed, neuroimaging work suggests that high-priority, task-relevant working memory representations are maintained via sustained

doi: 10.1111/nyas.13647

internal attention with persistent neural firing, while lower priority representations are maintained via alternative storage mechanisms (e.g., possibly synaptic weight changes).[42,43] To date, there have been few instances[26,44] of investigating attentional bias effects across these different states of working memory representations. Van Moorselaar *et al.*[26] found that memory for two items did not bias visual search as did memory for one item; however, if one of those two items was prioritized with a cue after encoding, only that item once again affected visual search. Greene *et al.*[44] manipulated item priority with an *n*-back task and concluded that the lower priority representation (the 1-back item) did not show attentional capture, while the higher priority item (2-back) did.

Recently, Dowd *et al.*[45,46] devised an approach to leverage sustained attentional capture effects in order to successfully decode the contents of working memory during a memory delay. The authors developed a method of "behavioral decoding" that aggregates small attentional biases in responses times (RTs) across a series of 12 visual search trials and uses machine learning to predict the identity of the working memory item responsible for the observed pattern of RTs. Here, we extend this approach to investigate whether items outside the focus of attention in working memory could be decoded from the pattern of attentional biases during visual search. Our task used retro-cues[47] to manipulate the priority of items such that they are either inside the focus (cued as relevant for the upcoming memory test) or outside the focus (cued as irrelevant for the first test, but potentially relevant for a second test) of attention.[40,48] Indeed, neuroimaging work has used this paradigm to emphasize the degree to which these representations are differentiable[49–52] and also observed that lower priority representations are latent but still have different response properties than permanently irrelevant items.[53–55]

In experiment 1, we replicated the findings of Dowd *et al.*[46] using a smaller sample of in-laboratory participants ($N = 16$ in laboratory versus $N = 100$ online). In experiment 2, we extended this procedure to include multiple memory items with a double retro-cue manipulation, and we investigated the existence of attentional capture from items that are inside versus outside the focus of attention. We used the patterns of RTs during two sets of visual searches from the first and second memory delays of each trial to decode the identities of the two memory items, and we also explored the time course of attentional capture across each delay period. Our data show that a single item in the focus of attention can bias visual attention toward matching perceptual items for quite a while (at least 24 s), whereas a memory item that is cued as (temporarily) task irrelevant shows only brief attentional capture effects (up to $\sim$ 3 s) while it is being removed to a lower priority state outside the focus of attention. At this point, it becomes *behaviorally silent* (i.e., it no longer systematically biases visual attention so as to be behaviorally decodable), unless it is cued as task relevant and re-enters the focus of attention, shortly after.

Experiment 1

The purpose of experiment 1 was to replicate the recent findings of Dowd *et al.*,[46] which demonstrated the ability to decode the contents of working memory from attentional biases on RTs across multiple visual search trials during a memory delay. Successful behavioral decoding of a working memory item indicates that it was systematically biasing visual attention. While their study used a large online participant sample from an Amazon Mechanical Turk population ($N = 100$), we sought to replicate their results with a smaller laboratory sample.

Materials and methods

Participants. Sixteen participants between the ages of 18 and 21 (five males) were recruited from the undergraduate student body of the University of Texas (UT) at Austin using the SONA system provided by the Department of Psychology. Participants were compensated with course credit for their participation in the 60-min experiment. Written informed consent was obtained in a manner approved by the UT Austin Institutional Review Board. Participants with overall memory accuracy below 2 SD of the mean were excluded ($n = 1$). No participants fell below 2 SD of the mean of visual search accuracy, and so no additional exclusions were made. All analyses were conducted on the remaining participants ($N = 15$).

Stimuli and procedure. Experiment 1 replicated the dual-task procedure from Dowd *et al.*[45,46] (Fig. 1A). Participants performed a simple memory recognition task including four colors. Each color was presented as the memory target 20 times,

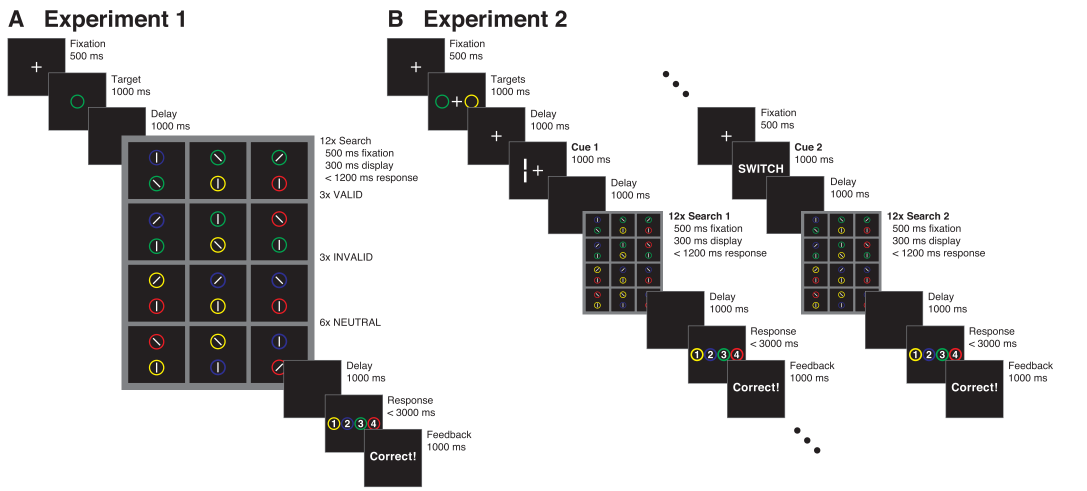

Figure 1. Task procedures. (A) Experiment 1 was a replication of an experiment performed by Dowd *et al.*[46] that required participants to hold a single item in working memory while performing an intervening sequence of 12 visual search trials. (B) Experiment 2 extended this paradigm to include two memory delays with retro-cues selecting which of two memory items would be tested after the delay (cue 1: a vertical line either left or right of center, which selected the item that had appeared on that side; cue 2: "STAY" or "SWITCH"). During each delay period, participants performed a randomized sequence of 12 visual search trials (identical to procedure in experiment 1).

shuffled across all blocks. Embedded within the delay period of each memory trial was a series of 12 visual search probes, where participants were asked to choose the direction (left or right) of a tilted line. Each search display consisted of one of 12 two-color combinations. This allowed, in every search set, for each of the four colors to surround the search target (three valid trials), to surround the distractor (three invalid trials), and to be absent from the search display (six neutral trials). The same combination of search displays was presented on each trial in randomized order, with the location (top/bottom) and tilt direction (left/right) of the target also randomized such that three targets appeared in each of the four combinations of top/bottom and left/right. Importantly, this randomization allowed for an equal amount of search targets presented within each of the four colors.

Participants practiced on one block of four memory trials before performing 10 blocks of eight trials in the experiment. For each memory trial, a central fixation cross was presented for 500 ms, followed by central presentation of the memory target for 1000 ms, with text above the target saying "Remember this color!" After a 1000-ms delay, the search task was presented, followed by another 1000-ms delay before the memory probe. The probe consisted of a four-alternative forced-choice task with four colored rings, one of each color, in which participants were to choose the color of the memory item that was presented at the beginning of the trial. The probe was presented for 3000 ms (until response), and feedback was presented ("Correct!" or "Incorrect.") for 1000 milliseconds.

Within a search set, each search probe began with a central fixation dot for 500 ms, followed by a search display with two potential search targets separated vertically (the center of each search target was 2.34° from center). One search display item contained the search target (a white line tilted 45° from vertical), and the other contained the distractor (a vertical line). Participants were instructed to respond as quickly as possible by choosing left or right as the direction of the tilted line. They could respond at any point during the 300-ms search display or following the 1200-ms blank delay period. Only negative feedback was provided ("Incorrect."). Stimuli were presented on a 21.5" iMac using Matlab 2014a and the Psychophysics Toolbox 3 extensions.[56] Responses to the search task were made using the left and right arrow keys with the right hand, and memory responses were made using the 1, 2, 3, and 4 keys with the left hand. All responses were collected using a standard USB keyboard. Stimuli were presented on a black background, and all text was presented in white. All fixations and search target lines were also white. Memory and search display rings were either red (RGB = 277, 2, 24), blue (48, 62, 152), green (95, 180, 46), or yellow (251, 189, 18). All rings (memory target, search displays, and memory probes) had a diameter of 2.34° (to outer edge) with a line thickness of 0.16°. All fixations had a diameter of 0.47°, and the search target lines had a diameter of 1.41° with a thickness of 0.16°.

Analysis

Multivariate behavioral classification. Multivariate classification methods were modeled after Dowd *et al.*[45,46] Before training the classification algorithm, RTs were preprocessed across all participants simultaneously. After removal of incorrect memory trials (8.1% of all trials), we replaced all RTs to inaccurate search probes (7.85% of all probes) with the mean RT of accurate search responses to effectively remove the influence of incorrect responses on the classifier. We also accounted for minor differences between colors by removing each color's deviation from the mean (M = 0.01 ms) to aid the normalization process. Next, we detrended all RTs run-wise, and finally all RTs were rescaled to [0:1]. We report results here without the addition of polynomial features based on the original 12 RTs on each trial, as was done by Dowd *et al.*[45] Our results were not affected by the addition of these features, and therefore they were excluded. After preprocessing, a multinomial logistic regression classifier with L2 regularization (implemented via sci-kit learn[57]) was trained and tested using a leave-one-subject-out cross-validation procedure. To gauge the statistical significance of classifier performance, we used a permutation test (10,000 iterations) to derive the empirical null distribution, in which we randomly shuffled the labels in our sample and recomputed classifier accuracy for each iteration. We reported P values as the proportion of shuffled scores that were higher than our observed score. To assess the population-level reliability of the classifier performance, we also ran a bootstrap resampling test (10,000 iterations) in which we resampled data from participants with replacement and recomputed classifier accuracy for each sample.

Time course of attentional capture. Search probes where the memory target surrounded the distractor in the search display are referred to as invalid probes (which should produce longer RTs), and valid probes are those where the memory target surrounded the search target (which should produce shorter RTs). We quantified the effect of attentional capture by calculating the difference metric d for each participant by subtracting the mean RT on valid probes from the mean RT on invalid probes. Before this difference calculation, the RTs were first preprocessed as described in the previous section, but without rescaling (to maintain temporal interpretation). If there was no attentional capture effect present, this metric would be close to zero. We evaluated the statistical significance of attentional capture across the entire search set using a one-sample t-test against zero (where each participant contributed a single difference metric d). To investigate the time course of the effect, we repeated this difference metric calculation, testing separately for each quartile of the search set with false discover rate (FDR) multiple-comparison correction applied. Each quartile was an amalgamation of nearby search responses (Q1: searches 1–3, Q2: 4–6, Q3: 7–9, and Q4: 10–12).

Results

Behavioral performance. Participants performed the tasks well, with a memory accuracy of 91.9 ± 1.6% SEM and a visual search accuracy of 92.2 ± 1.0% SEM across all trials.

Multivariate behavioral classification. The classifier was able to aggregate small deflections in RTs that arise from attentional capture effects from the item in working memory throughout the delay period. The memory target from experiment 1 was classified correctly across participants with an accuracy of 35.1% (Fig. 2). This accuracy was significant against shuffled-label permutation tests ($P < 0.001$). These results successfully replicate the findings of Dowd *et al.*,[46] who decoded working memory items with 36.5% accuracy under similar conditions.

Time course of attentional capture. To investigate the time course of attentional capture throughout the search set during the memory delay, we analyzed each quartile of the search set separately. There was an attentional capture effect when averaging the RTs across all search probes (Fig. 3A). This effect was also present at each quartile of the search set (Fig. 3B, all P values < 0.01). Notably, the effect of attentional capture

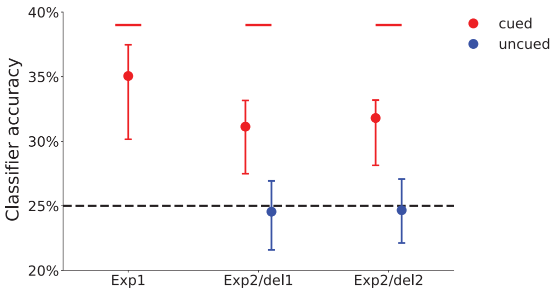

Figure 2. Behavioral decoding accuracy. Data with error bars represent the mean and 95% bootstrapped confidence intervals of classifier accuracy from leave-one-subject-out cross-validation analysis. Experiment 1 had only a single (cued) memory item, whereas experiment 2 had both a cued and uncued item in both the first delay period (del. 1) and the second delay period (del. 2). The dashed line represents theoretical chance accuracy (25%). Colored bars along the top indicate significant classification performance ($P < 0.001$).

decreased across the search set ($R^2 = 0.19$, $P < 0.001$).

Experiment 2

The purpose of experiment 2 was to use the behavioral-decoding procedure from experiment 1 to test for potential attentional bias effects from lower priority, temporarily irrelevant items outside the focus of attention in working memory.[49] We used a modified Sternberg task with two retro-cues[58] to experimentally manipulate the state of working memory items to be either inside or outside the focus of attention. Experiment 2 included two memory items per trial, with retro-cues selecting which item would be tested in the subsequent delay period. The 12-trial visual search task from experiment 1 was inserted into each delay period, and we used behavioral decoding to test whether an attentional bias could be detected from the relevant item in the focus, as well as from the irrelevant item outside the focus.

Materials and methods

Participants. Thirty participants between the ages of 18 and 34 (13 males) were recruited from the undergraduate student body of UT Austin using the SONA system provided by the Department of Psychology. Participants were compensated with course credit for their participation in the 90-min experiment. Written informed consent was obtained in a manner approved by the UT Austin Institutional Review Board. Participants were excluded if they had memory accuracy below 2 SD of the mean ($n = 3$), or visual search accuracy below 2 SD of the mean ($n = 2$). All analyses were conducted on the remaining participants ($N = 25$).

Stimuli and procedure. Experiment 2 modified the task from experiment 1 to include two memory delay periods with retro-cues (Fig. 1B). Retro-cue paradigms are used in working memory research as a tool to manipulate the priority of working memory items.[47] Two memory items were presented on each trial, followed by a first retro-cue that indicated (with 100% validity) which item would be tested by the first memory probe. After this test, a second retro-cue indicated which item would be tested by a second memory probe (again, with 100% validity). Specifically, the cue indicated whether the participant should stay with the same memory item or switch to the other item for this second test, with equal likelihood of receiving each cue across the experiment. This procedure produces a situation in which, during the first memory delay, the uncued item is only temporarily irrelevant, because it is potentially relevant for the second memory test. This item can be temporarily removed[59] from the focus of attention into an operationally defined low-priority state sometimes referred to as unattended,[49] secondary or accessory,[22] or activity-silent.[41] The uncued item during the second delay period, in contrast, is no longer needed and can be permanently removed from working memory.

Software, stimulus dimensions, and stage timings were identical to those of experiment 1 unless otherwise noted. Experiment 2 started with a practice block of four trials, followed by 12 blocks of eight memory trials each. After initial fixation, two memory targets were presented simultaneously on alternate sides of the central fixation cross (2.34° from center). Memory cues were presented such that each color combination (e.g., green and yellow) was presented equally often and balanced across the experiment as to whether a color was cued or uncued on the left or right. A fixation cross immediately followed encoding for 500 ms, and then a retro-cue appeared on either the left or right side of fixation for 1000 ms, indicating which of the two memory targets would be tested first. After a 1000-ms blank delay, the first search set began, followed by a delay and the first memory probe. Immediately afterward, a fixation cross was shown for 500 ms,

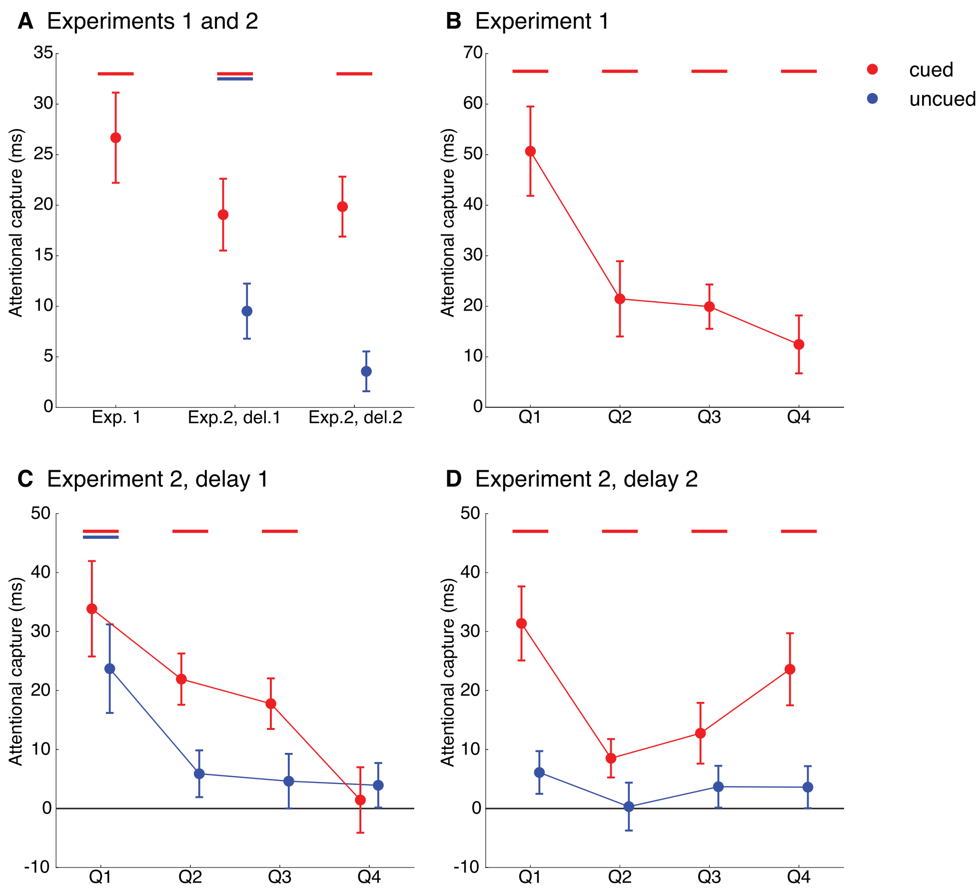

Figure 3. Attentional capture effects. RT differences between valid and invalid search probes are shown for both experiments 1 and 2, averaged across the entire delay period (A), and separated into quartiles (three searches) of the 12-trial visual search set (B, C, D). Data and error bars represent mean and SEM. Colored bars along the top indicate significant effects ($P < 0.05$ (FDR corrected)).

followed by a second retro-cue: centrally fixated text that read either "STAY" or "SWITCH." After a 1000-ms blank delay, the second search set was presented, and the trial was concluded with the second (and final) memory test of the trial.

Analysis

All analyses from experiment 1 were repeated for experiment 2. Each analysis was performed separately for the four conditions of delay (first/second) × cue (cued/uncued). We excluded trials if the response was incorrect on either the first or second memory probe (16.63% of trials).

Multivariate behavioral classification. The same classification procedures were used as described in experiment 1, with an added modification of using different labels for each trial depending on which item (cued or uncued) was being decoded. For example, the pattern of RTs from a given search set would be labeled *green* if the green ring was cued for that delay period and we were training a classifier to decode all cued items. However, the same data from that trial would be labeled *yellow* in a separate classifier if the yellow ring was uncued for that delay period and we were training a classifier to decode all uncued items. This cue-dependent labeling procedure allowed us to independently investigate cued and uncued items. During preprocessing of the visual search RTs, 6.86% of incorrect search probe RTs from the first delay period and 7.03% from the second delay period were replaced with mean search RTs. The mean RT deviation from the mean by color was 0.02 milliseconds.

Time course of attentional capture. To compute the difference measure d to quantify attentional capture (see experiment 1 for details), we wanted to account for the potentially counteractive bias effect of the other memory item when both memory targets were present on a search trial (e.g., when a green cued ring contained the distractor and a yellow uncued ring contained the target, or vice versa). Therefore, we removed any search trials that included both memory items (50% of all search trials) from the calculation of attentional capture (d) separately for each memory item.

Results

Behavioral performance. Overall memory accuracy was at 93.9 ± 1.0% SEM, and overall search accuracy was at 94.5 ± 0.7% SEM. Memory accuracy between the first delay (95.5 ± 0.8% SEM) and the second delay (92.3 ± 1.3% SEM) differed significantly ($t(24) = 3.46$, $P < 0.01$), as did accuracy between stay trials (95.5 ± 1.13% SEM) and switch trials (89.17 ± 1.8% SEM) after the second delay period ($t(24) = 4.01$, $P < 0.001$).

Multivariate behavioral classification. Classifier results are presented in Figure 2. During the first delay period, classification accuracy was 31.1% for the cued item, indicating significant decodability of this memory item in the focus of attention ($P < 0.001$). However, we were unable to behaviorally decode the uncued item (24.6%), putatively outside the focus of attention, during the first delay period ($P = 0.49$). In the second delay period, as in the first delay period, the cued item was successfully classified, with a mean accuracy of 31.8% ($P < 0.001$). Once again, however, the uncued item was undecodable during the second delay period (24.7%; $P = 0.48$).

Time course of attentional capture. Averaging across all search trials in the first delay period, we found evidence for significant attentional capture for both cued and uncued items (Fig. 3A). These effects were strongest during the beginning of the search set for both items (Fig. 3C). For the cued item, there was significant attentional capture during the first three quartiles (all P values < 0.001), but not in the final quartile ($P = 0.79$). For the uncued item, there was evidence of significant attentional capture in the first quartile ($P < 0.01$), similar in strength to the effect from the cued item ($P = 0.27$), but this effect was absent for the remainder of the delay period. In the second delay period (Fig. 3D), we observed significant attentional capture for the cued item averaged across the entire delay period, and also separately for each quartile. No significant attentional capture was observed for the uncued item during the second delay period.

Discussion

The primary goal of the current study was to investigate whether unattended items held in working memory exert a bias on visual attention. To test this, we combined a novel behavioral decoding method[45,46] that analyzes patterns of response time biases during visual search to decode the contents of working memory, with a retro-cueing paradigm[50,58] that manipulates whether an item is inside or outside the internal focus of attention in working memory. If a memory item (here, a colored circle) biases attention during visual search, that item (i.e., its color) should be decodable using this procedure. Our results indicate that, when two items are held in working memory, only the item inside the focus of attention consistently biases visual attention and is behaviorally decodable, whereas an item outside the focus does not bias attention sufficiently to be decoded. However, if that item is cued as relevant for a second memory test, it re-enters the focus and exerts an attentional bias, while the formerly relevant but now irrelevant item no longer does. These results provide further evidence for a functional distinction between items held inside and outside the focus of attention, and they demonstrate how preserving the richness of behavioral data by aggregating rather than averaging across responses can corroborate findings from neuroimaging studies of working memory.[49,50,53,60,61]

In experiment 1, our ability to behaviorally decode the identity of a single item held in working memory replicates recent findings from Dowd *et al.*[46] and is consistent with a large body of work pointing to an automatic capture of attention for items matching the contents of working memory.[1] Further analysis of our data revealed a decrease in this automatic attentional capture over the time course of the visual search trials (Fig. 3B), suggesting that there was some attenuation of the memory representation during the delay period. This attenuation is consistent with recent neuroimaging findings[62] showing the attenuation of the neural trace of working memory items and slowed response times across a filled delay period. This attenuation was even more pronounced when the delay-period task was more demanding on visual attention. These data suggest that both time and concurrent attentional demands influence the representational strength and the resulting attentional biases of a single task-relevant item in working memory. In experiment 2, we replicated the results from experiment 1 using a retro-cue to select, from a set of two memory items, a single item that was relevant for a first memory test. After this cue, the conditions were identical between the two experiments, except for the addition of a second memory item (in experiment 2) that could become relevant for testing after the first delay period. The cued item was behaviorally decodable during this delay period (Fig. 2) and showed a similar time course and attenuation of attentional bias throughout the delay as did a single memory item in experiment 1 (Fig. 3B and C). The uncued item, however, was not behaviorally decodable during this delay. We did find a brief attentional capture effect from this item early in the search set (across the first three search trials, lasting about 3 s), but this dissipated quickly and did not return. This suggests that the item was removed from the focus of attention, at which point it ceased to actively bias visual attention. The removal time for this item is consistent with behavioral[63] and neural estimates[49] of removing information from working memory.

Our finding of (albeit brief) attentional capture due to the uncued memory item contrasts with earlier investigations of multiple template representations in working memory that found an attentional bias limited to only one of two items.[23–26,44] They converge, however, with more recent results that show a bias from multiple items during increased loads.[27–31] These studies all have in common the investigation of attentional capture when multiple items are in working memory. But, importantly, there are often subtle task differences. For example, the finding that only a single representation can bias attention comes mostly from studies using a single memory item and an independently varying search template (Refs. 23 and 24 but not 26). In contrast, studies that involve two simultaneous working memory[29,30] or search items[27,28] typically find an effect of attentional capture of both items. When a search target changes from trial to trial, participants typically maintain an active representation, or template, of the search target in sensory regions,[32,64,65] similar to models of "sensory recruitment" storage for working memory items.[66] Yet, when the search target remains constant across trials, the target representation is thought to be stored in long-term memory.[33,67–69] Indeed, active search templates and working memory representations are largely similar, although they possibly differ in the effort required for their maintenance.[70,71] Our paradigm is most similar to previous studies that have explicitly assigned a higher priority to one of two memory items (and thus a lower priority to the other item).[25,26,44] Results from these studies are consistent with a model of visual working memory that proposes that attentional bias is limited to a single mnemonic representation.[22] While our behavioral decoding results support this model, our results showing a brief effect of attentional capture from low-priority items do not (at least on the surface). We address each of these past results separately.

Greene *et al.*[44] had participants respond to a 2-back memory task and found that the higher priority memory target that was 2-back biased attention, while the 1-back target did not. Stimuli used were similar to those in the current study (this has been suggested as the cause of previous conflicting results[72]). The different results likely can be attributed to different task paradigms: our retrocue task versus their *n*-back task. The *n*-back task is a significantly more demanding task requiring continual updating of working memory contents, and its convergent validity with established tests of working memory has been challenged.[73] The increased cognitive load of the *n*-back task may have obscured attentional biases that are captured in our retrocueing paradigm.

Results from van Loon *et al.*[25] come from a paradigm that explicitly manipulates priority, although there is no memory task per se, but rather both targets are retained for search (i.e., as attentional templates). Using metrics of eye movements rather than RTs, they found that a lower priority (search) target does not bias attention. The sensitivity of their metrics provides compelling results, yet there are no directly comparable investigations using memory tasks. Experiment 4 by van Moorselaar *et al.*[26] is largely similar to the current paradigm, and they found no attentional capture from a lower priority (memory) item. The task employed by van Moorselaar *et al.*[26] (as well as van Loon *et al.*[25]) used a continuous probe to promote the use of visual working memory. In contrast, our design was a four-choice categorical color decision, which might have allowed for subjects to recode one (or both) items into verbal working memory, as recoding is sometimes performed strategically.[51] Verbal working memory representations have been shown to bias attention,[4–7] and thus visual-to-verbal recoding (of either item) could underlie our finding of brief attention biases from both high- and low-priority memory items, while also remaining consistent with recent models of a single biasing representation in visual working memory.[22]

Returning to our discussion of experiment 2 in the present study, we now consider the effects from the second delay period. After the first delay period, a retro-cue selected which of the two memory items would be tested by the second (and final) memory test of that trial. The cued item—regardless of whether it was the same item or a different item than was selected by the first cue—was behaviorally decodable during this delay period. Not only did a task-relevant item in working memory maintain its automatic attentional bias across a second filled-delay period (for an additional $\sim$ 12 s), but a task-irrelevant item that lost its attentional bias during the first delay regained it and retained it throughout the second delay period. This reloading of the unattended item into the focus of attention, which has been shown to reinstate its active neural trace,[49–51] also reinstates its biasing influence on visual attention. The consequences of this flexible unloading and reloading of task-relevant items in working memory[74] support the idea of an overlap between internal and external attention, where not only does the demand on external attention influence the quality of representations in working memory,[62] but the focus of attention in working memory also influences the automatic biasing of external attention.

As found in the first delay period, the uncued item in the second delay period did not systematically and reliably bias visual attention, and it was therefore not behaviorally decodable. Interestingly, whereas we found an early, brief ($\sim$ 3 s) attentional bias for the uncued item during the first delay, there was no such effect during the second delay (Fig. 3D). There are two likely reasons for this. First, half of the trials had a second cue that matched the first cue, and thus the uncued item from the first delay (which was removed from the focus of attention and quickly lost any bias on visual attention) remained outside the focus during the second delay and never regained any attentional bias. Second, this difference may reflect the outcome of two different forms of removal after the first and second cues.[59] During the first delay period, two potentially relevant items in the (broad) focus of attention were separated by the retro-cue into one relevant item (that remained in the focus) and another item that was (temporarily) removed from the focus. This item was not forgotten and did not lose its binding to the trial context, because it could be reinstated into the focus if selected by the second retro-cue. Following the second retro-cue, the item that was uncued would never become relevant again, and was therefore a candidate for permanent removal from working memory.[59] If this process operates more quickly than temporary removal (see Ref. 59), this could diminish any attentional capture effects from the removed item during the second delay period.

Indeed, using retro-cues that indicated the uncued memory item could be dropped from memory, van Moorselaar *et al.*[75] found that attentional capture of the uncued (permanently irrelevant) item was absent at stimulus onset asynchronies (SOAs) as low as 100 milliseconds. The current work suggests that, when the item is cued as only temporarily irrelevant, a low-priority representation has an early effect of attentional capture. This contrast supports the separable processes of temporary and permanent removal.[59] However, other recent work found no attentional capture effects from low-priority items at SOAs ranging between 1000 and 1900 milliseconds.[25,26] The SOA in the current study (between the retro-cue and onset of the search task) was 1000 ms, and it is possible that longer SOAs might allow the process of temporary removal of the low-priority item to complete, thus eliminating its influence on visual attention. Future work investigating the effects of SOA on attentional capture from low-priority items would be helpful to clarify these inconsistencies.

We interpret our results according to models of working memory that describe multiple representational states and invoke some notion of the "focus of attention" to describe the most central, task-relevant component.[40,76] Alternatively, we could attempt to reconcile our results with a model of visual working memory[22] that describes "attentional templates" that guide visual search. The temporarily irrelevant, uncued items that briefly bias attention in our study likely map onto either secondary or accessory items in this model. However, it is important to again highlight that our paradigm did not require the visual storage of memory items per se. The memory items could easily (and are likely to) have been recoded into verbal forms,[51,77] for example, by remembering "red" instead of maintaining a visual representation of the red circle that was presented as the memory item. Many forms of mental representation beyond visual working memory have been observed to bias attention. Of particular relevance, biases can originate from verbal working memory.[4–7]

The absence of evidence does not imply evidence of absence. Without neuroimaging data, it is impossible to draw any conclusions about how the representation of temporarily irrelevant items in working memory might be stored in the brain. Even with brain data, the most we have been able to accomplish to date is to identify brief echoes of silent working memory items through exogenous stimulation with transcranial magnetic stimulation[53] or perceptual perturbations.[54,55] Despite the absence of a true smoking gun, it seems plausible that these items are indeed stored in an activity-silent fashion and retained for brief intervals by short-term changes in synaptic plasticity[42,43] (but see Ref. 78). In summary, our present findings suggest that working memory representations outside the focus of attention, which are known to become (reversibly) neurally silent, also become (reversibly) behaviorally silent in terms of not exerting automatic biases on visual attention.

Acknowledgments

R.M. and J.L.P. conceived the study, designed the experiment, and wrote the paper. R.M. collected and analyzed data under the supervision of J.L.P. The authors would like to thank Katlyn Hedgpeth for contributions to data collection and Kimber Mallett for contributions to figure design. This work was supported by funds from the University of Texas System STARs Program awarded to J.L.P.

Competing interests

All authors declare no competing interests regarding this research.

References

1. Soto, D., J. Hodsoll, P. Rotshtein, *et al.* 2008. Automatic guidance of attention from working memory. *Trends Cogn. Sci.* **12:** 342–348.
2. Downing, P.E. 2000. Interactions between visual working memory and selective attention. *Psychol. Sci.* **11:** 467–473.
3. Soto, D., D. Heinke, G.W. Humphreys, *et al.* 2005. Early, involuntary top-down guidance of attention from working memory. *J. Exp. Psychol. Hum. Percept. Perform.* **31:** 248–261.
4. Kawashima, T. & E. Matsumoto. 2017. Cognitive control of attentional guidance by visual and verbal working memory representations. *Jpn. Psychol. Res.* **59:** 49–57.
5. Soto, D. & G.W. Humphreys. 2007. Automatic guidance of visual attention from verbal working memory. *J. Exp. Psychol. Hum. Percept. Perform.* **33:** 730–737.
6. Mannan, S.K., C. Kennard, D. Potter, *et al.* 2010. Early oculomotor capture by new onsets driven by the contents of working memory. *Vision Res.* **50:** 1590–1597.
7. Soto, D., P. Rotshtein, J. Hodsoll, *et al.* 2012. Common and distinct neural regions for the guidance of selection by visuoverbal information held in memory: converging evidence from fMRI and rTMS. *Hum. Brain Mapp.* **33:** 105–120.
8. Fan, J.E. & N.B. Turk-Browne. 2016. Incidental biasing of attention from visual long-term memory. *J. Exp. Psychol. Learn. Mem. Cogn.* **42:** 970–977.
9. Rosen, M.L., C.E. Stern & D.C. Somers. 2014. Long-term memory guidance of visuospatial attention in a change-detection paradigm. *Front. Psychol.* **5:** 266.
10. Summerfield, J.J., J. Lepsien, D.R. Gitelman, *et al.* 2006. Orienting attention based on long-term memory experience. *Neuron* **49:** 905–916.
11. Woodman, G.F. & M.M. Chun. 2006. The role of working memory and long-term memory in visual search. *Vis. Cogn.* **14:** 808–830.
12. Moores, E., L. Laiti & L. Chelazzi. 2003. Associative knowledge controls deployment of visual selective attention. *Nat. Neurosci.* **6:** 182–189.
13. Sun, S.Z., J. Shen, M. Shaw, *et al.* 2015. Automatic capture of attention by conceptually generated working memory templates. *Atten. Percept. Psychophys.* **77:** 1841–1847.
14. Johnson, J.S., G.F. Woodman, E. Braun, *et al.* 2007. Implicit memory influences the allocation of attention in visual cortex. *Psychon. Bull. Rev.* **14:** 834–839.
15. Anderson, B.A., P.A. Laurent & S. Yantis. 2011. Value-driven attentional capture. *Proc. Natl. Acad. Sci. USA* **108:** 10367–10371.
16. Carlisle, N.B. & G.F. Woodman. 2011. When memory is not enough: electrophysiological evidence for goal-dependent use of working memory representations in guiding visual attention. *J. Cogn. Neurosci.* **23:** 2650–2664.
17. Woodman, G.F. & S.J. Luck. 2007. Do the contents of visual working memory automatically influence attentional selection during visual search? *J. Exp. Psychol. Hum. Percept. Perform.* **33:** 363–377.
18. Awh, E. & J. Jonides. 2001. Overlapping mechanisms of attention and spatial working memory. *Trends Cogn. Sci.* **5:** 119–126.

19. Chun, M.M., J.D. Golomb & N.B. Turk-Browne. 2011. A taxonomy of external and internal attention. *Annu. Rev. Psychol.* **62:** 73–101.
20. Gazzaley, A. & A.C. Nobre. 2012. Top-down modulation: bridging selective attention and working memory. *Trends Cogn. Sci.* **16:** 129–135.
21. Kiyonaga, A. & T. Egner. 2013. Working memory as internal attention: toward an integrative account of internal and external selection processes. *Psychon. Bull. Rev.* **20:** 228–242.
22. Olivers, C.N., J. Peters, R. Houtkamp, *et al.* 2011. Different states in visual working memory: when it guides attention and when it does not. *Trends Cogn. Sci.* **15:** 327–334.
23. Downing, P. & C. Dodds. 2004. Competition in visual working memory for control of search. *Vis. Cogn.* **11:** 689–703.
24. Houtkamp R. & P.R. Roelfsema. 2006. The effect of items in working memory on the deployment of attention and the eyes during visual search. *J. Exp. Psychol. Hum. Percept. Perform.* **32:** 423–442.
25. van Loon, A.M., K. Olmos-Solis & C.N.L. Olivers. 2017. Subtle eye movement metrics reveal task-relevant representations prior to visual search. *J. Vis.* **17:** 13.
26. van Moorselaar, D., J. Theeuwes & C.N. Olivers. 2014. In competition for the attentional template: can multiple items within visual working memory guide attention? *J. Exp. Psychol. Hum. Percept. Perform.* **40:** 1450–1464.
27. Beck, V.M., A. Hollingworth & S.J. Luck. 2012. Simultaneous control of attention by multiple working memory representations. *Psychol. Sci.* **23:** 887–898.
28. Beck, V.M. & A. Hollingworth. 2017. Competition in saccade target selection reveals attentional guidance by simultaneously active working memory representations. *J. Exp. Psychol. Hum. Percept. Perform.* **43:** 225–230.
29. Hollingworth, A. & V.M. Beck. 2016. Memory-based attention capture when multiple items are maintained in visual working memory. *J. Exp. Psychol. Hum. Percept. Perform.* **42:** 911–917.
30. Chen, Y. & F. Du. 2017. Two visual working memory representations simultaneously control attention. *Sci. Rep.* **7:** 6107.
31. Bahle, B., V.M. Beck & A. Hollingworth. 2018. The architecture of interaction between visual working memory and visual attention. *J. Exp. Psychol. Hum. Percept. Perform.* In press.
32. Wolfe, J.M. 1994. Guided search 2.0 a revised model of visual search. *Psychon. Bull. Rev.* **1:** 202–238.
33. Grubert, A., N.B. Carlisle & M. Eimer. 2016. The control of single-color and multiple-color visual search by attentional templates in working memory and in long-term memory. *J. Cogn. Neurosci.* **28:** 1947–1963.
34. Cowan, N. 2010. The magical mystery four: how is working memory capacity limited, and why? *Curr. Dir. Psychol. Sci.* **19:** 51–57.
35. Ma, W.J., M. Husain & P.M. Bays. 2014. Changing concepts of working memory. *Nat. Neurosci.* **17:** 347–356.
36. Oberauer, K., S. Farrell, C. Jarrold, *et al.* 2016. What limits working memory capacity? *Psychol. Bull.* **142:** 758–799.
37. Cowan, N. 1988. Evolving conceptions of memory storage, selective attention, and their mutual constraints within the human information-processing system. *Psychol. Bull.* **104:** 163–191.
38. McElree, B. 1998. Attended and non-attended states in working memory: accessing categorized structures. *J. Mem. Lang.* **38:** 225–252.
39. Nee, D.E. & J. Jonides. 2013. Trisecting representational states in short-term memory. *Front. Hum. Neurosci.* **7:** 796.
40. Oberauer, K. 2002. Access to information in working memory: exploring the focus of attention. *J. Exp. Psychol. Learn. Mem. Cogn.* **28:** 411–421.
41. Stokes M.G. 2015. 'Activity-silent'working memory in prefrontal cortex: a dynamic coding framework. *Trends Cogn. Sci.* **19:** 394–405.
42. Mongillo, G., O. Barak & M. Tsodyks. 2008. Synaptic theory of working memory. *Science* **319:** 1543–1546.
43. Barak, O. & M. Tsodyks. 2014. Working models of working memory. *Curr. Opin. Neurobiol.* **25:** 20–24.
44. Greene, C.M., K. Kennedy & D. Soto. 2015. Dynamic states in working memory modulate guidance of visual attention: evidence from an *n*-back paradigm. *Vis. Cogn.* **23:** 546–560.
45. Dowd, E.W., J.M. Pearson & T. Egner. 2015. Mind-reading without the scanner: behavioural decoding of working memory content. *Vis. Cogn.* **23:** 862–866.
46. Dowd, E.W., J.M. Pearson & T. Egner. 2017. Decoding working memory content from attentional biases. *Psychon. Bull. Rev.* **24:** 1252–1260.
47. Souza, A.S. & K. Oberauer. 2016. In search of the focus of attention in working memory: 13 years of the retro-cue effect. *Atten. Percept. Psychophys.* **78:** 1839–1860.
48. van Moorselaar, D., C.N. Olivers, J. Theeuwes, *et al.* 2015. Forgotten but not gone: retro-cue costs and benefits in a double-cueing paradigm suggest multiple states in visual short-term memory. *J. Exp. Psychol. Learn. Mem. Cogn.* **41:** 1755–1763.
49. LaRocque, J.J., J.A. Lewis-Peacock, A.T. Drysdale, *et al.* 2013. Decoding attended information in short-term memory: an EEG study. *J. Cogn. Neurosci.* **25:** 127–142.
50. Lewis-Peacock, J.A., A.T. Drysdale, K. Oberauer, *et al.* 2012. Neural evidence for a distinction between short-term memory and the focus of attention. *J. Cogn. Neurosci.* **24:** 61–79.
51. Lewis-Peacock, J.A., A.T. Drysdale & B.R. Postle. 2015. Neural evidence for the flexible control of mental representations. *Cereb. Cortex* **25:** 3303–3313.
52. Sprague, T.C., E.F. Ester & J.T. Serences. 2016. Restoring latent visual working memory representations in human cortex. *Neuron* **91:** 694–707.
53. Rose, N.S., J.J. LaRocque, A.C. Riggall, *et al.* 2016. Reactivation of latent working memories with transcranial magnetic stimulation. *Science* **354:** 1136–1139.
54. Wolff, M.J., J. Ding, N.E. Myers, *et al.* 2015. Revealing hidden states in visual working memory using electroencephalography. *Front. Syst. Neurosci.* **9:** 123.
55. Wolff, M.J., J. Jochim, E.G. Akyürek, *et al.* 2017. Dynamic hidden states underlying working-memory-guided behavior. *Nat. Neurosci.* **20:** 864–871.
56. Brainard, D.H. 1997. The psychophysics toolbox. *Spat. Vis.* **10:** 433–436.
57. Pedregosa, F., G. Varoquaux, A. Gramfort, *et al.* 2011. Scikit-learn: machine learning in Python. *J. Mach. Learn. Res.* **12:** 2825–2830.
58. Oberauer, K. 2005. Control of the contents of working memory—a comparison of two paradigms and two

age groups. *J. Exp. Psychol. Learn. Mem. Cogn.* **31:** 714–728.
59. Lewis-Peacock, J.A., Y. Kessler & K. Oberauer. 2018. The removal of information from working memory. *Ann. N.Y. Acad. Sci.* **XXXX:** XX–XX.
60. Lewis-Peacock, J.A. & B.R. Postle. 2012. Decoding the internal focus of attention. *Neuropsychologia* **50:** 470–478.
61. van Moorselaar, D., J.J. Foster, D.W. Sutterer, *et al.* 2018. Spatially selective alpha oscillations reveal moment-by-moment trade-offs between working memory and attention. *J. Cogn. Neurosci.* **30:** 256–266.
62. Kiyonaga, A., E.W. Dowd & T. Egner. 2017. Neural representation of working memory content is modulated by visual attentional demand. *J. Cogn. Neurosci.* **29:** 2011–2024.
63. Oberauer, K. 2001. Removing irrelevant information from working memory: a cognitive aging study with the modified Sternberg task. *J. Exp. Psychol. Learn. Mem. Cogn.* **27:** 948–957.
64. Stokes, M., R. Thompson, A.C. Nobre, *et al.* 2009. Shape-specific preparatory activity mediates attention to targets in human visual cortex. *Proc. Natl. Acad. Sci. USA* **106:** 19569–19574.
65. Stokes, M.G. 2011. Top-down visual activity underlying VSTM and preparatory attention. *Neuropsychologia* **49:** 1425–1427.
66. D'esposito, M. & B.R. Postle. 2015. The cognitive neuroscience of working memory. *Annu. Rev. Psychol.* **66:** 115–142.
67. Carlisle N.B., J.T. Arita, D. Pardo, *et al.* 2011. Attentional templates in visual working memory. *J. Neurosci.* **31:** 9315–9322.
68. Reinhart, R.M. & G.F. Woodman. 2015. Enhancing long-term memory with stimulation tunes visual attention in one trial. *Proc. Natl. Acad. Sci. USA* **112:** 625–630.
69. Woodman, G.F., S.J. Luck & J.D. Schall. 2007. The role of working memory representations in the control of attention. *Cereb. Cortex* **17:** i118–i124.
70. Olivers, C.N.L. & M. Eimer. 2011. On the difference between working memory and attentional set. *Neuropsychologia* **49:** 1553–1558.
71. Gunseli, E., M. Meeter & C.N. Olivers. 2014. Is a search template an ordinary working memory? Comparing electrophysiological markers of working memory maintenance for visual search and recognition. *Neuropsychologia* **60:** 29–38.
72. Olivers, C.N. 2009. What drives memory-driven attentional capture? The effects of memory type, display type, and search type. *J. Exp. Psychol. Hum. Percept. Perform.* **35:** 1275–1291.
73. Kane, M.J., A.R. Conway, T.K. Miura, *et al.* 2007. Working memory, attention control, and the *N*-back task: a question of construct validity. *J. Exp. Psychol. Learn. Mem. Cogn.* **33:** 615–622.
74. Rerko, L., A.S. Souza & K. Oberauer. 2014. Retro-cue benefits in working memory without sustained focal attention. *Mem. Cognit.* **42:** 712–728.
75. van Moorselaar, D., E. Battistoni, J. Theeuwes, *et al.* 2015. Rapid influences of cued visual memories on attentional guidance. *Ann. N.Y. Acad. Sci.* **1339:** 1–10.
76. Cowan, N. 1995. *Attention and Memory: An Integrated Framework.* Oxford University Press.
77. Wickens, D.D. 1973. Some characteristics of word encoding. *Mem. Cognit.* **1:** 485–490.
78. Schneegans, S. & P.M. Bays. 2017. Restoration of fMRI decodability does not imply latent working memory states. *J. Cogn. Neurosci.* **29:** 1977–1994.