PART 2

ATTENTION TO MEMORY
Abstract

Retrospectively cueing an item retained in visual working memory during maintenance is known to improve its retention. However, studies have provided conflicting results regarding the costs of such retro-cues for the non-cued items, leading to different theories on the mechanisms behind visual working memory maintenance and retro-cueing. Here we tested an alternative explanation of the conflicting results regarding retro-cue costs, namely that they are at least partly caused by differences in retro-cue reliability. We manipulated the ratio of valid cue trials to invalid cue trials within blocks. We used a continuous report procedure that allowed fitting a model that provided recall probability and precision estimates for memory representations. Reconciling previous contradictory findings, benefits for valid cues were observed in all conditions, but invalid cueing costs were found only when the retro-cue had a high reliability (i.e., was 80% valid) but not when it had a lower reliability (i.e., 50% valid). This was found for both the recall probability and the precision of visual working memory representations. Our results suggest that the cognitive mechanisms underlying retro-cue effects are strategically adjusted by participants depending on perceived retro-cue reliability.
Chapter 5

Introduction

Visual working memory (VWM) is the cognitive system in which a limited amount of visual information can be briefly maintained and manipulated. Attention interacts with many stages of VWM processing, including encoding (Posner, 1980; Schmidt, Vogel, Woodman, & Luck, 2002; Vogel, Luck, & Shapiro, 1998), maintenance (Awh & Jonides, 2001; Awh, Jonides, & Reuter-Lorenz, 1998; Munneke, Heslenfeld, & Theeuwes, 2010), and retrieval (Theeuwes, Kramer, & Irwin, 2011). One way to look at this interaction is through the use of retro-cues. These are typically spatial cues presented during maintenance that point out one of the memory items, which then becomes particularly likely to be tested. It has been shown that such retro-cues result in improved memory performance compared to trials without a retro-cue (Griffin & Nobre, 2003; Landman et al., 2003; Lepsien et al., 2005). This retro-cue benefit has been claimed to reflect (a) the reallocation of attentional resources within memory, resulting in the protection of the cued representation against decay and interference (Protection Hypothesis, Makovski & Jiang, 2007; Makovski et al., 2008; Matsukura et al., 2007; Pertzov et al., 2013; van Moorselaar, Gunseli, Theeuwes, & Olivers, under review-a), (b) removing non-cued items from memory, therefore presumably reducing the inter-item interference and competition for resources (Removal Hypothesis, Kuo et al., 2012; Souza, Rerko, & Oberauer, 2014; Williams & Woodman, 2012), (c) carrying the cued item to a more robust or ‘prioritized’ state during maintenance without altering non-cued items (Prioritization During Maintenance, Myers et al., 2014; Rerko & Oberauer, 2013; Souza et al., 2014), or (d) prioritize the cued representation during retrieval without affecting maintenance per se (Prioritization During Retrieval, Astle et al., 2012; Nobre, Griffin, & Rao, 2007).

Although all these hypotheses predict a benefit for the cued representation, they differ in their assumptions regarding the costs for non-cued representations. The Protection and Removal hypotheses predict that retro-cueing benefits for the cued representation should be accompanied by costs for non-cued representations because they involve reallocation of resources away from non-cued items towards the cued one. The Prioritization During Maintenance and Prioritization During Retrieval hypotheses, on the other hand, predict no such costs because they explain the cue benefits by a change in the status of the cued item without any change for non-cued items. Previous studies using retro-cues have so far produced conflicting results. Some have observed costs in recognition or recall performance when a non-cued representation is probed (Matsukura et al., 2007; Pertzov et al., 2013), while others have observed no such costs in these so-called invalid trials (Landman et al., 2003; Lepsien & Nobre, 2007; Rerko & Oberauer, 2013). The studies that did not observe any invalidity costs used a double-cueing paradigm in which an invalid cue is followed by a valid cue. This may make participants more hesitant to drop an item after the first cue. However, Matsukura et al. (2007)
observed a cost of invalid retro-cueing while using a double-cueing paradigm. Therefore, we believe double-cueing itself cannot account for the inconsistency across findings regarding invalidity costs. Yet again, other studies have observed costs only in reaction time and not accuracy, which has been interpreted as support for the Prioritization During Retrieval hypothesis (Astle et al., 2012). The lack of costs for non-cued items in recognition accuracy has gained further theoretical significance because it has been taken as evidence for the idea that VWM maintenance does not require any active rehearsal via attention (for a similar argument, see Hollingworth & Maxcey-Richard, 2013; Rerko & Oberauer, 2013). Thus, knowing whether retro-cues result in costs for non-cued representations is theoretically important for understanding the mechanisms behind VWM maintenance, as well as those behind retro-cueing.

Although we cannot nor wish to exclude the possibility of different factors playing a role, the present study investigated a factor that may at least partially explain the inconsistency in cue-related costs, namely the reliability of the cue. A similar argument has been made by Williams and Woodman (2012) in the context of directed forgetting cues. Cue reliability can be operationally defined as the ratio between valid and invalid trials. Typically, studies that failed to observe a cost in invalid trials had relatively low cue reliabilities (50% valid, Landman et al., 2003; 50% valid, Lepsien & Nobre, 2007; 66.6% valid, Rerko & Oberauer, 2013) than those that observed a cost (80% valid, Astle et al., 2012; 75% valid, Matsukura et al., 2007; 70% valid, Pertzov et al., 2013). It may thus be the case that when a cue has a high reliability (e.g. a high valid to invalid trial ratio), participants devote most of their attentional resources to the cued representation (i.e. Protection) and remove the non-cued items from memory (i.e. Removal) since there is very little chance of being tested on them. On the other hand, when a cue has a low reliability (e.g. a low valid to invalid trial ratio), participants may keep on maintaining non-cued representations in anticipation of potentially being tested on them. In this case, they may merely prioritize the cued item during maintenance and/or retrieval, without costs to the non-cued representations. This straightforward hypothesis might account for the inconsistencies in the literature regarding the costs of invalid retro-cues.

We tested the cue reliability hypothesis by manipulating the validity of the retro-cue. Participants were asked to remember orientations of four bars and then recall the orientation of one probe bar. On some trials, during the maintenance interval a probabilistic retro-cue was presented: It pointed to the subsequently probed item in 80% of trials on some blocks (80% validity), and in 50% of trials on other blocks (50% validity).

---

4 Note that some studies defined validity ratios calculated as the number of trials valid/(valid+neutral+invalid) instead of valid/(valid+invalid) . In the present study we used the latter since we did not expect trials without a cue to affect the subjective evaluation of the reliability of the cue.
On the remaining, invalid trials it pointed to one of the items that was not subsequently probed. Participants were informed about these validity ratios before each block. Furthermore, rather than employing the often-used change detection / recognition task, in which observers can only provide a discrete same/different judgment, we used a continuous-recall procedure that provides a more sensitive measure of maintenance of VWM representations that provides a degree of quality of recall performance, rather than reducing it to a binary decision (Bays, Catalao, & Husain, 2009; Wilken & Ma, 2004; Zhang & Luck, 2008). This measure also enabled fitting a model that estimates the recall probability of the target, non-target (i.e. non-probed) representations, and also the precision of memory. To foreshadow the main findings, retro-cues improved the recall probability and precision of the target. Importantly, both the benefits of valid retro-cues and the costs of invalid retro-cues were greater when the cue was highly reliable (i.e. 80% valid in comparison to 50% valid), to the extent that an invalidity cost was absent for probability and precision in the low reliability condition.

Method

Experimental Procedures

Twenty-two healthy volunteers participated in the experiment for course credit or monetary compensation. For twelve of the participants, we also took EEG recordings for another study. Their behavioral performance was no different from the rest. Two participants were excluded from analysis due to low performance (see Analysis). The study was conducted in accordance with the Declaration of Helsinki and was approved by the faculty’s Ethical Committee. Written informed consent was obtained.
The Reliability of Retro-Cues Determines the Fate of Non-cued Memory Representations

Figure 5.1. The experimental procedure in the present experiment. The retro-cue was a fixation circle with a quarter filled with either red or green to point one of the memory representations. Similarly, the test probe was indicated by a white quarter filling. In this example, participants needed to report the orientation of the bar presented on bottom-right corner, and the retro-cue is valid since it pointed in the same position. There were also trials in which the retro-cue was invalidly pointing to a bar that was not going to be tested. In neutral (i.e. no-cue) trials, the fixation dot remained on the screen during the retro-cue duration. During test, participants had to rotate the orientation of the bar to match that in their memory using the mouse.

The procedure is shown in Figure 1. The memory display consisted of four black oriented bars (2.08° x 0.25° visual angle) located equidistantly on an imaginary circle of radius 3.50°, and was presented for 350 ms. The orientation of each bar was chosen at random with the restriction that bars within the same trial differed by at least 10°. The test display was presented 1550 ms after the offset of the memory display (1650 ms for the 12 participants for whom the EEG was also recorded). It contained a randomly oriented bar and a cue pointing to the location of the probe representation that were both presented at the center of the screen. This probe cue was the same as the fixation circle except that a quarter (90°) of it was filled white. Participants were asked to indicate the precise orientation of the bar at the probed location by rotating the probe bar using the mouse. After a mouse response was made, the correct orientation was indicated by a central white bar for 100 ms. The inter-trial interval was 800 ms for ten participants, and a jittered interval between 1200-1600 ms for the other twelve.
On retro-cue trials, after the memory display, there was a maintenance interval of either 550 ms (for ten participants) or 650 ms (for 12 participants), followed by the presentation of the retro-cue display for 100 ms. The retro-cue was the same as the probe cue except the fill color was either red, 27.08 Cd/m², or green, 24.10 Cd/m², depending on the reliability condition (order counterbalanced). For the initial practice phase where the cue was 100% valid, the retro-cue fill color was orange (53.46 Cd/m²). Following the retro-cue, there was a second maintenance interval of 900 ms. In no-cue trials, the black fixation circle remained on the screen during the whole maintenance interval without any changes to it. The timing of the test display was matched for retro-cue and no-cue trials.

In the high reliability condition, the cue was 80% valid and in the low reliability condition it was 50% valid. The reliability conditions were blocked. In each experimental block, 25% of the trials were neutral - that is, there was no cue. Each validity condition (i.e. valid, neutral, and invalid) was randomly intermixed within each block. Before each reliability condition, participants were informed about the validity ratio of the cues (as was also indicated by the color of the cue), and they performed a practice session of 25 trials to get used to this particular validity ratio. Moreover, at the beginning of the experiment, there was an initial practice session with a 100% valid cue that contained 20 trials (25 for 12 of the participants), to make participants familiar with the cue. In total, there were 560 trials (600 for 12 of the participants). In order to have a reasonable number of invalid cue trials, there were more blocks of the high reliable cue condition than of the low reliable cue condition. Respectively for valid, neutral, and invalid trials, there were 216, 90, and 54 trials for 80% valid cue, and 75, 75, and 50 trials (90, 90, and 60 for 12 of the participants) for 50% valid cue. The main constraint was to have at least 50 trials per condition for a reliable model fit (see, http://www.paulbays.com/code/JV10/index.php). At the end of each block, participants received feedback on block average and grand average memory deviation values.

Analysis

Deviation scores on the memory test were calculated as the average difference (i.e. error) between the original orientation of the probed memory bar and the orientation of the response. The precision was calculated, per condition, as the inverse of the standard deviation of the error in subjects’ responses (Bays et al., 2009). The deviation scores were entered into a model to calculate the probability of recall for the target and non-target VWM representations (Bays et al., 2009). Two participants were excluded from further analysis due to low performance: one had a target recall probability barely above chance level (i.e. above the chance of reporting any of the four orientations) in the 50% valid no-cue condition (a fitted recall probability of .26 and an average deviation of 37.1 degrees), and the other had a recall probability of .51 (and an average deviation of 34.8) in 80%
valid no-cue condition that was almost the same probability of reporting one of the non-probed items (i.e. .49). Moreover, these target recall probabilities were 2.5 times the standard deviation lower than the overall mean for the given condition. The important results were the same when these participants were included. Raw deviation, target recall probability, non-target recall probability and precision for each condition were entered into a repeated-measures ANOVA with the factors of reliability (80% valid vs. 50% valid) and validity (valid, neutral, and invalid). Contingent on a significant reliability x validity interaction, these were followed up by separate ANOVAs testing for validity benefits (i.e. the difference between neutral and valid trials) and invalidity costs (i.e. the difference between invalid and neutral trials). Where necessary, p-values were adjusted for sphericity violations using the Greenhouse–Geisser epsilon correction on degrees of freedom (Jennings & Wood, 1976). To test whether the validity benefits and invalidity costs were different than zero, one-sample t-tests were used.

**Results**

We first tested whether the two reliability conditions differed in how participants performed on neutral trials. Any differences on neutral trials would have suggested that altering cue reliability would have changed the way participants approached the whole task. For all measures described below (i.e. raw deviation, precision, recall probability for the target, and the recall probability for the non-target), the performance on neutral trials did not differ between 80% valid and 50% valid conditions (all ts <1.00, ps>.330).

**Raw Deviation**

Next, we looked at the effect of cue validity and reliability on raw deviations from the target orientation. Figure 2A shows the distribution of errors for each condition, at which each data point represents the frequency of errors for bins of 15 degrees of deviations (Pertzov & Husain, 2013). There was no main effect of reliability on deviation, $F(1, 19) = 2.10, p =.163, \eta_p^2=.10$, but there was one of validity, $F(2, 38) = 10.23, p <.001, \eta_p^2=.35$. Importantly, there was a reliability x validity interaction, $F(2, 38) = 26.15, p <.001, \eta_p^2=.58$. Planned comparisons showed that both the validity benefit (i.e., the difference between valid and neutral trials), $t(19) = 2.32, p =.032$, and the invalidity cost (i.e. the difference between invalid and neutral trials), $t(19) = 2.64, p =.023$, were larger for the 80% valid condition compared to the 50% valid condition. Both the validity benefit and invalidity cost were present (i.e. significantly different than zero) for both reliability conditions ($ts>2.37, ps<.028$).
Precision

The average precision in each condition is shown in Figure 2B. Again there was a main effect of validity on precision, $F(2, 38) = 74.54, p < .001, \eta^2_p = .80$, none of reliability, $F(1, 19) = .04, p = .838, \eta^2_p = .01$, and a reliability x validity interaction, $F(2, 38) = 16.84, p < .001, \eta^2_p = .47$. Both the validity benefit, $t(19) = 2.31, p = .032$, and the invalidity cost, $t(19) = 3.68, p = .002$, were greater in the 80% than in the 50% condition. The invalidity cost was significant for 80% valid condition, $t(19) = 4.67, p < .001$, but not for 50% valid condition, $t(19) = .56, p = .580$. The validity benefit was greater than zero in both conditions, ($t > 5.35, p < .001$).

Recall probability for the target

The average recall probability in each condition is shown in Figure 2C. There were main effects of reliability on probability, $F(1, 19) = 5.06, p = .036, \eta^2_p = .21$, and of validity, $F(2, 38) = 13.75, p < .001, \eta^2_p = .42$, and a reliability x validity interaction, $F(2, 38) = 5.91, p = .018, \eta^2_p = .24$. Planned comparisons showed that the validity benefit did not differ between the 80% valid and the 50% valid conditions, $t(19) = 1.41, p = .174$, while the invalidity cost was larger for the 80% valid condition compared to the 50% valid condition, $t(19) = 2.12, p = .047$. The validity benefit was present (i.e. significantly different from zero) for both conditions ($t > 2.16, p < .044$), while the invalidity cost was present in the 80% valid condition, $t(19) = 2.45, p = .024$, but not the 50% valid condition, $t(19) = .26, p = .795$. 
Figure 5.2. (A) Distribution of errors relative to the target (i.e. probed) orientation for 50% valid (left panel) and 80% valid (right panel) conditions. (B) Precision for the target, (C) recall probability estimate for the target, and (D) recall probability estimate for a non-target for each condition. The invalid, neutral and valid trials are shown in red, blue, and green respectively. (E) Distribution of errors on invalid trials relative to non-target orientations. The error bars represent standard mean errors for standardized data (i.e. corrected for between-subjects variance, Cousineau, 2005). The ns, ‘*’ and ‘**’ represent p>.05, p<.05 and p<.005 respectively.
Recall probability for non-targets

The average probability of recalling a non-probed item in each condition is shown in Figure 2D. There was a main effect of validity, $F(2, 38) = 7.09, p = .013, \eta^2_p = .27$, and none of reliability on this probability, $F(1, 19) = 1.61, p = .220, \eta^2_p = .08$. Again, there was a reliability x validity interaction, $F(2, 38) = 4.66, p = .039, \eta^2_p = .20$. Planned comparisons showed that the validity benefit (in terms of a lower likelihood of recalling a non-probed item on valid than on neutral trials) was not different for 80% valid and 50% valid conditions, $t(19) = .12, p = .904$, even though it was significant, in post-hoc tests, only in the 80% valid, $t(19) = 2.15, p = .045$, and not in the 50% valid condition, $t(19) = 1.08, p = .294$. The invalidity cost, meanwhile, was higher for 80% valid condition compared to 50% valid condition, $t(19) = 2.09, p = .051$. The probability of reporting a non-probed item was greater for invalid trials compared to neutral trials only in the 80% valid condition, $t(19) = 2.36, p = .029$, but not in the 50% valid condition, $t(19) = .99, p = .336$.

In order to test if the high probability of reporting a non-target item was driven by reporting the cued non-target or any of the (non-cued) non-targets, we compared the error distributions around the orientation of the cued non-target and non-cued non-targets (see Figure 2E). The distribution of responses around the cued non-target on 80% valid trials was somewhat steeper compared to the cued non-target on 50% valid trials and compared to non-cued non-targets in both reliability conditions (although the difference in percentage of errors at -7.5 and 7.5 degrees did not reach significance, $t < 1.73, ps > .100$). This leaves open the possibility, although statistically not supported, that the higher non-target recall probability on 80% invalid trials in comparison to other conditions is due to recalling the cued non-target on a greater proportion of trials than recalling any other non-target.

Discussion

The findings support the idea that the degree of retro-cue effects on recall performance depends on the reliability of the cue. The cost of invalid cueing was minor for raw deviation, and altogether absent for precision and recall probability estimates when the cue was relatively unreliable (i.e. 50% valid), while there was still a clear benefit for valid cues. When the cue was more reliable (80% valid), benefits were larger, and now costs were also present. Furthermore, on invalid trials, the likelihood of mistakenly reporting a non-probed item during test was higher when the cue was more reliable. These results suggest that how participants implement the retro-cue to the memory task is, at least partly, under strategic control: When the cue has low reliability, observers prioritize the cued item for maintenance and/or retrieval without letting go of the non-cued items (Prioritization During Maintenance and Prioritization During Retrieval) probably in anticipation of the still quite likely event of being tested on one of the non-cued items.
a result invalid cueing costs are at most minor. In contrast, when the cue is highly reliable, in addition to prioritization, attentional and/or memory resources are disengaged from non-cued items during maintenance (Protection and Removal), which leads to a high invalidity cost when a non-cued item is probed. Retro-cue effects thus seem to be in line with the Prioritization During Maintenance or Prioritization During Retrieval hypotheses when cue validity is low, but in line with the Protection or the Removal hypotheses when cue validity is high.

Cue reliability may not be the only contributing factor in determining invalidity costs. For example, Astle et al. (2012) found that invalid cues had a cost on recognition accuracy only when memory set size exceeded the VWM capacity limit (i.e. eight), but not for set sizes within the VWM capacity limit (i.e. two and four, but see van Moorselaar, Olivers, Theeuwes, Lamme, & Sligte, under review-b), despite the fact that their cue was 80% valid. Using the same set size and the same cue reliability, in the present study we observed a cost of invalid cueing. Our study and that of Astle et al. (2012) differ in the test used to measure memory performance. We believe that the continuous report procedure used in the present study is a more sensitive memory measure and therefore might reveal differences in performance that are less likely to be detected with the discrete same/different judgment task that was used by Astle et al. (2012), because it provides a measure of how good the response is for each trial instead of reducing the response to a binary decision (Wilken & Ma, 2004). Consistent with this claim, in the present study, the effects of retro-cueing were more pronounced for precision compared to the recall probability of items. Nevertheless, we cannot exclude the possibility that invalidity costs might be smaller, although not completely absent, for smaller set sizes even with a continuous recall measure since the possibility of being tested on a particular non-cued item is higher, and also maintenance is less demanding for smaller set sizes. Both of these factors make it less beneficial to redistribute attentional/memory resources when the set size is small.

Notwithstanding the role of set size, our findings suggest that some of the inconsistency in results on invalidity costs is due to differences in the reliability of the retro-cue (for a similar argument for directed forgetting cues, see Williams & Woodman, 2012). Thus, the absence of an invalidity cost in Rerko and Oberauer (2013) may merely reflect a lack of attentional redistribution due to the low reliability of the retro-cue, rather than the absence of a role of attention in VWM maintenance. Our results suggest that attentional redistribution is performed mostly for highly reliable cues (as in 80% valid cue condition in the present study) and that without being attended, VWM representations are vulnerable to interference and/or decay – consistent with earlier claims (Astle et al., 2012; Makovski & Jiang, 2007; Makovski et al., 2008; Matsukura et al., 2007; Pertzov et al., 2013; van Moorselaar et al., under review-a). Another possibility is that non-cued
items are actively removed from memory when cues are highly reliable, as this would also result in significant invalidity costs (Kuo et al., 2012; Souza et al., 2014; Williams & Woodman, 2012). Considering that previous research has provided support for both mechanisms, we believe that they may both occur. On some trials non-cued items may be actively removed from memory, whereas on other trials they are attended less, and therefore more vulnerable to interference. The important conclusion we want to make is that either mechanism is more likely to be implemented when the cue is highly reliable.

Regardless of these exact mechanisms, our findings point to a dissociation between how retro-cues affect the cued and the non-cued items. While one is attended, the other may be unattended but is not necessarily dropped (Rerko & Oberauer, 2013). Instead whether a non-cued item is maintained or not may be a separate decision. Such a dissociation is consistent with several models promoting a distinction between memory items that are in the current focus (“template”) and other VWM representations that are held prospectively, or “on reserve” (LaRocque et al., 2013; LaRocque et al., 2014; Oberauer, 2002; Olivers et al., 2011; Rerko & Oberauer, 2013; van Moorselaar et al., under review-b; van Moorselaar, Theeuwes, & Olivers, In Press; Zokaei et al., 2014a). These two types of VWM representation may have different mechanisms of maintenance, which operate more or less independently: 1. Task-relevant (here cued) representations are carried into a prioritized template status (which may also prioritize them for retrieval) regardless of the cue reliability, as there is little to lose by doing so. In line with this Berryhill, Richmond, Shay, and Olson (2012) demonstrated the presence of a validity benefit even when the cue was only informative on a minority of trials. 2. Currently irrelevant non-cued items are held via a more passive accessory storage, and observers may decide to remove these depending on the perceived reliability of the cue - that is depending on whether they see a potential future use for them.

In short, present results show that how retro-cues affect recall performance depends on the reliability of the cues. When highly reliable, retro-cues resulted in major invalidity costs and larger validity benefits as compared to low reliable retro-cues that resulted in minor invalidity costs and smaller validity benefits). Thus, cue reliability will have to be considered before drawing any conclusions from research using probabilistic retro-cues.