

Orienting attention to locations in mental representations

Duncan Edward Astle · Jennifer Summerfield ·
Ivan Griffin · Anna Christina Nobre

Published online: 5 October 2011
© Psychonomic Society, Inc. 2011

Abstract Many cognitive processes depend on our ability to hold information in mind, often well beyond the offset of the original sensory input. The capacity of this visual short-term memory (VSTM) is limited to around three to four items. Recent research has demonstrated that the content of VSTM can be modulated by top-down attentional biases. This has been demonstrated using retrodictive spatial cues, termed “retro-cues,” which orient subjects’ attention to spatial locations within VSTM. In the present article, we tested whether the use of these cues is modulated by memory load and cue delay. There are a number of important conclusions: (1) Top-down biases can operate on very brief iconic traces as well as on older VSTM representations (Exp. 1). (2) When operating within capacity, subjects use the cue to prioritise where they initiate their memory search, rather than to discard uncued items (Exps. 2 and 3). (3) When capacity is exceeded, there is little benefit to top-down biasing relative to a neutral condition; however, unattended items are lost, with there being a substantial cost of invalid spatial cueing (Exp. 3). (4) These costs and benefits of orienting spatial attention differ across iconic memory and VSTM representations when VSTM capacity is exceeded (Exp. 4).

Keywords VSTM · Visual short-term memory · Visual working memory · Iconic memory · Spatial attention · Partial report

D. E. Astle
Department of Psychology, Royal Holloway,
University of London,
London, UK

D. E. Astle · J. Summerfield · I. Griffin · A. C. Nobre (✉)
Department of Experimental Psychology, University of Oxford,
South Parks Road,
Oxford OX1 3UD, UK
e-mail: kia.nobre@psy.ox.ac.uk

Attention enables us to bias the processing of incoming perceptual information, favouring those aspects of the input most relevant to the task at hand (Posner, 1980). Recent studies have demonstrated that attention, in addition to operating upon perceptual representations, can bias the contents of visual short-term memory (VSTM; Astle, Scerif, Kuo, & Nobre, 2009; Griffin & Nobre, 2003; Landman, Spekreijse, & Lamme, 2003; Lepsien, Griffin, Devlin, & Nobre, 2005; Lepsien & Nobre, 2006; Matsukura, Luck, & Vecera, 2007; Makovski, Sussman, & Jiang, 2008; Sligte, Scholte, & Lamme, 2008, Sligte, Scholte, & Lamme, 2009; Sligte, Vandenbrouke, Scholte, & Lamme, 2010).

Whilst much is known about how we bias aspects of our perceptual representations, how the contents of memory can be biased is only beginning to be explored. Recent studies have used attention-directing retro-cues to demonstrate that the contents of VSTM are available for top-down biasing. Retro-cues are presented after the to-be-remembered arrays of items and predict which location or item from within the remembered array will be relevant to a subsequent decision based on the memory representation, such as judging whether a probe item was contained within the array. Retro-cues give subjects time to orient their attention to the cued location prior to the probe stimulus appearing. Retro-cues, therefore, differ from postcues, which directly act as probes for accessing a specific aspect of the remembered array. Postcues give no time to orient; subjects are instructed to respond immediately. Sligte et al. (2008) compared performance with a postcue and with a retro-cue. They demonstrated that capacity estimates more than doubled when subjects received a retro-cue rather than a postcue, even though the interval between the array and the cue was the same in both cases, and the actual delay until the probe item was longer in the retro-cue condition. The authors argued for the existence of a store of many “fragile”

VSTM representations, which were wiped by the onset of the postcue, but which were boosted by the attention resulting from the retro-cue, meaning that they could survive the onset of the probe.

The present set of experiments aimed to extend our understanding of how these retro-cues operate on memory representations. Work so far has ruled out a series of uninteresting artefactual explanations for some of these effects—for instance, speed–accuracy trade-offs (Griffin & Nobre, 2003; Lepsien et al., 2005), response biases (Griffin & Nobre, 2003), eye movements (Griffin & Nobre, 2003; Matsukura et al., 2007), or articulation (Makovski, Shim, & Jiang, 2006; Makovski et al., 2008; Matsukura et al., 2007). Other, more interesting possibilities remain: The retro-cue might enable subjects to enhance the active maintenance of cued items and/or to suppress the active maintenance of uncued items, retro-cues might enable subjects to insulate particular items from decay or interference, or retro-cues might enable subjects to prioritise where they start their memory search. (It is important to note that these possibilities need not be mutually exclusive.)

In the present set of experiments, we used valid, invalid, and neutral retro-cues. As in previous experiments, the valid retro-cues enabled us to explore the advantage of committing attention to a stored item, relative to not committing attention to any one item (i.e., relative to a neutral-cue baseline; see, e.g., Griffin & Nobre, 2003). Furthermore, using invalid cues enabled us to test the fate of those items that were uncued, by occasionally probing those items on a subset of trials. One possibility was that subjects would use the cue to select a single item and discard uncued items from memory (maybe by allowing them to degrade; see Matsukura et al., 2007), meaning that invalid cues would have a catastrophic effect on accuracy measures: The uncued items would be lost and become unavailable for reinspection at probe onset. By contrast, if subjects merely used the retro-cue to prioritise where to start their VSTM search at probe onset, rather than to actually discard some items, invalid cues would then have relatively little effect on accuracy measures. Uncued items would still be maintained in VSTM, just as well as the cued items, but subjects would initiate their VSTM search in the wrong location; whilst their responses would be slowed, little accuracy cost, relative to neutral cues, should be observed. In short, manipulating cue validity in this way would enable us to ask important questions as to how these attentional mechanisms operate on the contents of memory to bring about validity benefits and/or invalidity costs. This distinction is similar to the one between “protection” and “prioritisation” given by Matsukura et al. (2007), with the difference being that we are suggesting that a pure prioritisation account, in which cued and uncued items are equally available at probe onset, would predict cue validity effects on RTs but not on accuracy.

There are various ways of exploring the effect of retro-cueing on the uncued items. Matsukura et al. (2007) used a double-cue paradigm. Subjects were retro-cued to a subset of memory array shapes, and then, on a minority of trials, retro-cued to the other half of the shapes—that is, those that had previously been uncued. This initial cue acted like an invalid cue, cueing subjects to a subset of items that would not be needed at probe onset. Accordingly, accuracy rates were much lower when subjects following these alternating double-cue trials. Having discounted various alternative accounts, the authors concluded that subjects do not merely use the retro-cues to prioritise where they start searching their memory; rather, the effect of the retro-cues is that uncued items are lost from VSTM. This could occur either because subjects simply stop maintaining/insulating them from decay or because they are actively suppressed to remove them from VSTM. Either way, it is clear that the retro-cue does more than prioritise where to initiate memory search. An alternative way of testing the availability of uncued items is to use the single-cue paradigm and, on a minority of trials, to probe the uncued items. This approach has the advantage of removing the additional passage of time from conditions for assessing the state of uncued items. Matsukura et al. demonstrated that the cost of an invalid single cue is very similar to the cost of an initial invalid cue in the double-cue paradigm. If subjects use retro-cues to protect certain items from being lost (or to lose uncued items intentionally), subjects should perform poorly when one subsequently cues (double cue) or probes (invalid single cue) those previously uncued items. However, the relative costs and benefits of retro-cueing, and thus how subjects use the retro-cue, may change as a function of various factors, such as cue delay and memory load. That is, subjects may use the cues differently depending on these factors. The modulation of the benefits and costs of retro-cueing by these two factors is the principle question of the four experiments herein.

The effect of cue latency on retro-cue costs and benefits

Sperling (1960) used a partial-report paradigm to demonstrate that more information is “available” from large arrays than can be reported under usual task conditions. When subjects performed a whole-report on an array of visual letters—being asked to report as many items as possible—they were reliably able to retrieve up to around four items. However, when they were selectively probed about the contents of only one of the rows according to an auditory cue following the array, the estimate of the number of items maintained over a very brief period was much higher. When the cue was delivered at 150 or 300 ms post array offset, the estimated number of items available was, in some cases, double that in the whole-report condition (the effect was

subsequently replicated with visual cues: Averbach & Coriell, 1961). However, when this cue was delayed by 1,000 ms, the partial-report benefit was lost, with performance being equivalent to that in the whole-report condition. Sperling concluded that a substantial number of items are held in memory immediately after the offset of the array; however, within the first second of this maintenance delay, all but around four of these items are lost (Averbach & Coriell, 1961; Dick, 1974; Sperling, 1960). Of course, the specific array–cue intervals inevitably provide underestimates of the longevity of different types of memory traces, since some time would undoubtedly be required for processing and interpreting the auditory cues, meaning that the availability of these extra items might persist for slightly longer than Sperling had estimated.

In light of this functional distinction proposed between very brief, *iconic* memory (Neisser, 1967) and the more durable VSTM, we asked whether top-down attentional biases might work differently on these types of traces to enhance memory-based performance. We asked whether the relative benefits and costs of retro-cues might differ depending on whether they operate on the iconic memory (IM) trace or on a more durable VSTM representation. This question is explored in Experiments 1, 2, and 4.

The effect of memory load on retro-cue costs and benefits

By contrast with IM, the capacity of VSTM is strictly limited (Bays, Catalao, & Husain, 2009; Cowan, 2001; Vogel & Machizawa, 2004; Vogel, Woodman, & Luck, 2001; Zhang & Luck, 2008). For example, people's failure to notice large changes in visual scenes unless attention is specifically directed to them is usually used to infer some limit to VSTM capacity (O'Regan, Rensink, & Clark, 1999; Rensink, O'Regan, & Clark, 1997). In a series of experiments, Luck and colleagues, as well as others, have suggested that VSTM can only store approximately four integrated visual items, irrespective of the number of constituent features (Lee & Chun, 2001; Luck & Vogel, 1997; Vogel et al., 2001).

It is possible, therefore, that the relative costs and benefits of spatial retro-cues differ across loads. This is explored in Experiments 3 and 4. For instance, when the array size is comfortably within capacity (≤ 4 items), subjects may use the spatial retro-cue to prioritise where they initiate their memory search, but maintain all of the items nonetheless, producing little accuracy cost of invalid cueing. Alternatively, when the number of to-be-remembered items exceeds capacity, subjects may use the cue to prioritise which items they maintain, discarding uncued items and thereby reducing memory load. In short, the way in which a

retro-cue is used may differ depending on whether or not VSTM capacity is exceeded.

Given what has already been demonstrated (see, e.g., Sligte et al., 2008; Sperling, 1960), we might expect the two factors of cue latency and memory load to interact. When cue delays are short, a larger-capacity IM system is tapped; when cue delays are long, a smaller-capacity VSTM system is tapped. In our final experiment, we tested this proposition by manipulating both cue SOA and memory load in the same design, to explore the potentially interactive effects of these factors on subjects' use of the retro-cue.

Across the experiments, we report three different dependent measures: RT is an index of the speed with which items can be retrieved correctly, but it is insensitive to the number of items available for retrieval. Given that many of our questions pertain to the availability of items for retrieval, rather than the speed with which they can be retrieved, measures of accuracy might be more important. We present two measures of accuracy: (1) d' indexes accuracy, taking into account any response biases, and (2) K is also based on subjects' retrieval accuracy, but it takes into account the number of items in the original array (Cowan, 2001).

Experiment 1: Benefits of spatial orienting to items held in iconic or visual short-term memory

Most previous experiments have investigated the advantages gained from retrospective attentional cueing when the contents of VSTM, upon which those retro-cues operate, have been stored for between 1 and 2.5 s (Astle, Nobre, & Scerif, *in press*; Griffin & Nobre, 2003; Nobre, Griffin, & Rao, 2008). In one recent study (Sligte et al., 2008), retro-cue SOA varied up to 5 s, and even at these long retention intervals the retro-cue provided a benefit relative to performance on postcue trials. However, this previous study did not compare retro-cue performance to that on a neutral baseline, so it could not be determined whether the *benefit* of retro-cues changed as the contents of VSTM decayed, relative to a situation in which attention is never committed to any single item. We tested the temporal limits of retrospective attentional cueing, by varying the retention period between the onset of the to-be-remembered array of items and the retrospective cue, and we used the neutral cues as a baseline condition. We varied the SOA between the array and retro-cue using seven intervals between 150 and 9,600 ms. This also enabled us to address the additional question of whether retro-cues provide similar degrees of benefit when they operate on the contents of iconic memory (IM; e.g., 150 ms) or VSTM (e.g., 2,040 ms). We were keen to test whether there was continuity in the benefits derived from retrospective attentional cueing across these delays. One might predict a decreasing relative benefit of

retro-cues with increasing SOA—as the contents of VSTM decay, so that the stored items are not as available for attentional selection. Alternatively, one might predict that at longer durations VSTM items could become too fragile for retrieval via standard serial search mechanisms (Sternberg, 1966), but still available for attentional selection. Moreover, this attentional selection (following a retro-cue) might provide the stability needed for successful retrieval (Slight et al., 2008). Were that the case, we might find that the *relative benefit* of valid retro-cues would increase with increasing decay.

In addition to this effect of decay, one might predict that the benefit gained from a valid retro-cue would be different when operating on the contents of IM, by comparison with VSTM, because of the differing amounts and natures of the information available for attentional selection in these two forms of storage (Averbach & Coriell, 1961; Sperling, 1960).

Materials and method

Subjects The experimental methods in this and in all subsequent experiments had ethical approval from the Central University Research Ethics Committee at the University of Oxford. The subjects were healthy paid volunteers from the University of Oxford community of students and researchers. All had normal or corrected-to-normal visual acuity and were right-handed. Twelve subjects (age range 20–31 years; 7 females, 5 males) took part in this experiment.

Stimuli and task The task is illustrated in Fig. 1. Subjects viewed arrays of four differently coloured crosses, followed by an informative or neutral cue, and they made a delayed

decision about the colour of the items in the array. At the end of each trial sequence, a probe stimulus appeared; the subjects’ task was to decide whether the probe stimulus had been in the preceding memory array. The time interval between the onset of the array and cue (stimulus onset asynchrony, or SOA) varied between 150 and 9,600 ms, with the interval doubling successively across seven intervals (i.e., 150, 300, 600 ms, etc.).

Each trial contained the same sequence of events. A square (side length 0.8°, central cue stimulus) appeared at the centre of the screen for a random interval between 600 and 900 ms. An array of four differently coloured crosses was then presented for 100 ms. The crosses were any four of the following colours: red, blue, green, yellow, orange, cyan, pink, or grey. Each cross was 0.8° of visual angle in size and was centred at 3° horizontal and 3° vertical eccentricity. There was then another interval, between 50 and 9,500 ms (150- and 9,600-ms SOAs), after which either an informative or a neutral cue was presented. The cue appeared at one of seven possible SOAs, with equal probabilities and in a randomised fashion: 150, 300, 600, 1,200, 2,400, 4,800, or 9,600 ms. Following the cue, after a random interval between 500 and 1,000 ms, a coloured cross (any one of the eight possible colours; *probe stimulus*) appeared at the centre of the screen for 100 ms. Subjects responded by pressing the left button of a response box with their right-hand index finger if the probe stimulus did appear in the array (“yes” response) and the right button with their right-hand middle finger if the probe stimulus did not appear in the array (“no” response).

An informative cue consisted of two adjacent sides of the central square brightening (forming an arrow). Informative cues occurred in 67% of trials and accurately predicted (with 100% validity) the location where the probe stimulus had occurred if it had been present in the array. Neutral cues occurred in 33% of trials. They consisted of the whole square brightening and gave no spatial information about the likely location of the probe.

The probabilities of correct “yes” and “no” responses were equal; that is, 50% of the time the probe stimulus was present in the array, and 50% of the time it was not. This was true for both informative and neutral trials, at all SOAs. Both informative and neutral trials, at all SOAs, occurred interspersed in a random order throughout the experiment.

There were 588 trials (392 informative, 196 neutral; i.e., 84 at each SOA). Of the informative trials, 196 were target-present valid trials (probe appeared in array at the cued location), and 196 were target-absent trials. Of the neutral trials, 98 included a probe that was in the array, and 98 included a probe that was not in the array. Within informative and neutral trials, the seven SOAs (150, 300, 600, 1,200, 2,400, 4,800, and 9,600 ms) and the four cue directions (top left, top right, bottom right, and bottom left)

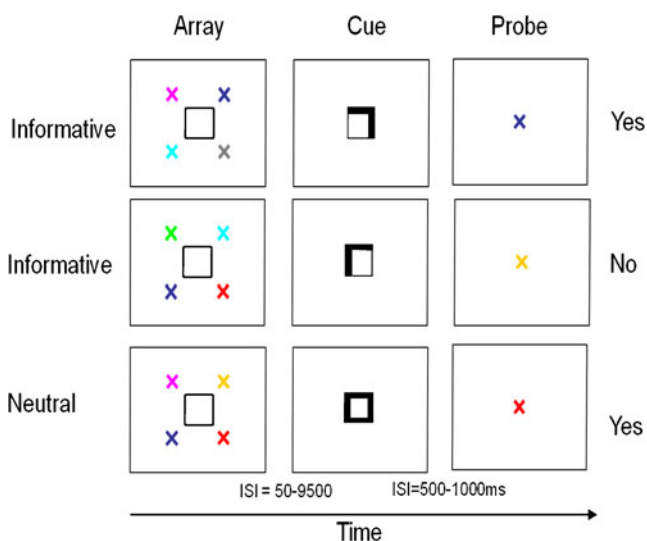


Fig. 1 A trial order schematic (Exp. 1) for two trials with informative spatial retro-cues and one neutral trial. The upper two trial schematics show probe-present trials, and the lower schematic shows a probe-absent trial

occurred equally frequently. There were 14 blocks of trials in the experiment, plus 1 additional practice block at the beginning. In all of our analyses in which the assumption of sphericity was violated, the Greenhouse–Geisser correction for nonsphericity was applied.

Procedures Subjects were comfortably seated in a dimly illuminated room, facing a computer monitor placed 100 cm in front of them. They were informed about the relationship between the cue, the array, and the probe stimuli. They were asked to maintain fixation on a small cross that was continuously present at the centre of the monitor. They were instructed to respond as quickly as possible following probe stimulus onset, whilst avoiding mistakes. No feedback was given during the experiment.

Results

Accuracy The accuracy scores were converted into d' scores (the normalised proportion of correct hits minus the normalised proportion of false alarms). This provides an advantage over analysing the raw accuracy scores in that it controls for any general response bias that might vary as a function of cue type (neutral vs. valid). These were submitted to a two-way ANOVA, with the within-subjects factors of cue type and SOA. The results can be seen in Fig. 2. There was a significant effect of SOA on d' scores

[$F(6, 66) = 3.008, p = .012$], with a significant trend for decreasing d' scores with increasing SOA [linear contrast for SOA; $F(1, 11) = 14.487, p = .003$]. There was a significant effect of cue [$F(1, 11) = 10.906, p = .007$], with d' scores being higher for valid (2.44 ± 0.23 [SE]; throughout the text, data points are represented as $M \pm SE$) than for invalid (2.05 ± 0.31) cues. There was no significant interaction between cue and SOA [$F(6, 66) = 0.283, p = .943$].

Reaction times Only correct trials were used in the RT analyses. The RT data were submitted to a three-way repeated measures ANOVA, with the within-subjects factors of SOA, cue type, and presence (probe present or absent). This revealed a significant main effect of cue [$F(1, 11) = 20.174, p = .001$], reflecting quicker responses on informative (525 ± 24 ms) than on neutral trials (630 ± 20 ms). A main effect of target presence [$F(1, 11) = 43.172, p < .001$] indicated quicker responses to probes that had been present in the array (578 ± 27 ms) than to those that had not (676 ± 22 ms). There was also a significant main effect of SOA [$F(6, 66) = 11.674, p < .001$]. In a follow-up analysis, we established that this effect of SOA did not stem from a significant linear trend of increasing RTs with increasing SOA (as we might have expected, given the d' scores) [$F(1, 11) = 1.164, p = .304$], but rather from a significant quadratic trend [$F(1, 11) = 35.867, p < .001$]. The pattern of RTs formed a U shape. Post-hoc contrasts revealed that RTs differed significantly between 600 and 1,200 ms [$F(1, 11) = 26.98, p < .001$], between 2,400 and 4,800 ms [$F(1, 11) = 10.92, p = .007$], and between 4,800 and 9,600 ms [$F(1, 11) = 18.45, p = .001$]. In short, RTs significantly decreased between 600 and 1,200 ms, and then increased again after 2,400 ms, resulting in a significant quadratic trend. There was also a significant interaction between probe presence and SOA [$F(6, 66) = 2.315, p = .043$]. This was because the effect of SOA was greatest on probe-present trials [$F(6, 66) = 10.571, p < .001$], relative to the effect of SOA on probe-absent trials [$F(6, 66) = 4.707, p < .001$]. The RTs from probe-present trials can also be seen in Fig. 2.

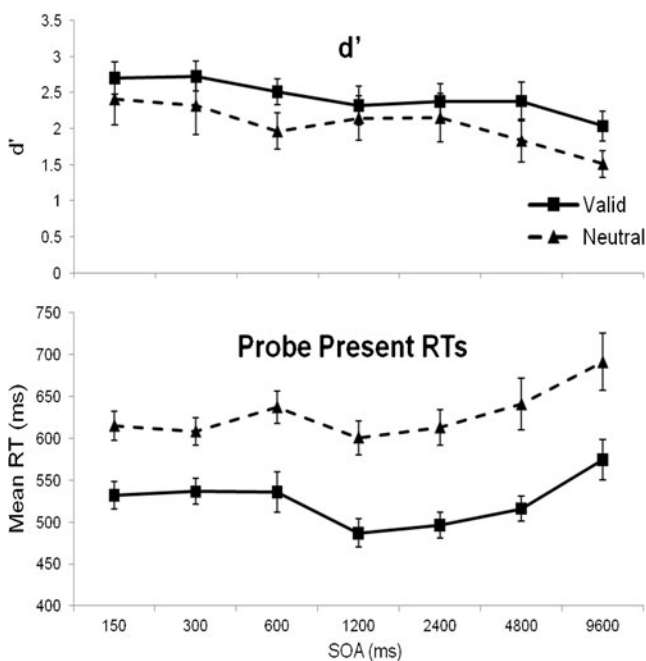


Fig. 2 Results from Experiment 1. The upper panel shows d' scores for valid and neutral trials across the SOAs. The lower panel shows reaction times (RTs) for the same comparison. In all cases, the error bars show the standard errors of the means

Discussion

Experiment 1 explored the relative benefits of retrodictive cues when they operated across SOAs within the IM and VSTM ranges, even up to ~ 10 s after the offset of the array. To our knowledge, this is the first time that such an analysis has been done. Behavioural benefits of orienting attention to internal representations were evident over a range of time periods, including SOAs at which representations are thought to be held in IM, as well as at much longer SOAs at which

representations are held in VSTM. The benefit derived from retro-cues relative to neutral cues was largely consistent across the various SOAs. Despite decreasing d' scores with increasing SOA, the relative d' benefit of valid retro-cues did not mirror this pattern (or show the opposite pattern).

Recent studies have shown that retro-cueing is effective in boosting decayed “fragile” VSTM representations (Sligte et al., 2008), but that the benefit of a retro-cue, relative to performance on a postcue trial, decreases with increasing SOA up to 4 s. After 4 s, performance in both the retro- and postcue conditions appears to decline. Thus, our data replicate the finding of decreasing performance with valid retro-cues with increasing decay, but because we compared this decreasing performance with a neutral condition, we could see that the rates of performance decline were roughly equivalent across these two trial types, and did not occur because retro-cues become gradually less effective per se.

It might appear that our results contradict previous demonstrations that cues are most effective when operating on an IM representation (e.g., Sperling, 1960). However, these enhanced benefits have been apparent when subjects are shown supra-VSTM-capacity-size arrays and are presented with postcues, which act as an imperative prompt for retrieval. These classic studies have been used to infer the existence of some iconic store with a capacity that exceeds that of VSTM. In our case, the number of items did not exceed VSTM capacity, and our retro-cues, unlike postcues, enabled preparation for a subsequent imperative stimulus. Thus, our demonstration that the cues operated equally well across the different SOAs does not necessarily contradict previous findings. Of course, if more than four items are available at the iconic time range, we might expect the factors of cue SOA and memory load to interact, with the cue being most effect when operating on supra-VSTM-capacity IM (see Exp. 4).

To summarise briefly, the results of Experiment 1 extend the previous observations of retro-cue benefits (e.g., Griffin & Nobre, 2003; Lepsien et al., 2005; Sligte et al., 2008) by demonstrating that the benefit of committing attention to a particular item, relative to not committing it to any one item, is evident over a very broad time range; searching through representations held as brief IM traces or as more durable VSTM representations can be biased by spatial attention, even if they are stored for ~10 s. However, the effect that this cueing has on the remaining uncued items remains unclear, and this question is addressed in Experiment 2.

Experiment 2: Benefits and costs of orienting spatial attention to items in iconic and visual short-term memory

It is well documented that valid retro-cues confer a substantial retrieval benefit (e.g., Griffin & Nobre, 2003;

Landman et al., 2003; Makovski et al., 2008; Nobre et al., 2008), and Experiment 1 explored the temporal constraints of this benefit. However, why is retro-cueing so advantageous? Using invalid cues is one potential way of exploring this question (see also Matsukura et al., 2007). One possibility is that subjects use the cue to *prioritise* their memory search. If this is the case, the occasional invalid cue would slow subjects down, since their search would be initiated in the wrong location; this would, however, have relatively little effect on accuracy scores, since the uncued items would still be available for reinspection. Alternatively, subjects might use the retro-cue to *discard* uncued items, essentially reducing the memory load to a single item. Were this the case, subjects' accuracy scores would be detrimentally affected on occasional invalid trials; the uncued items would no longer be available, and subjects would therefore infer that this was a target-absent trial. In short, using invalid trials would offer us the opportunity to explore what happens to the uncued items when subjects orient their attention within memory.

In Experiment 2, as with the cueing benefit in Experiment 1, we explored how the cost of invalid cueing changed across SOAs. Whilst valid retro-cueing benefits might be relatively stable across SOAs, invalid retro-cueing costs might not. For instance, if subjects have been invested in maintaining all four array items for a long period of time (e.g., ~10 s), it might be easier simply to keep maintaining the items but to prioritise the location at which they start their memory search. By contrast, if they have only been maintaining the items for a very brief period of time (e.g., 150 ms), subjects might then opt to use the cue to discard the uncued items and save themselves the effort of maintaining them all. In short, how subjects use the retro-cue, and thus what happens to the uncued array items, might change depending on how late the retro-cue is presented. This was explored in Experiment 2.

Materials and method

Unless stated otherwise, the methods used in Experiment 2 were identical to those in Experiment 1.

Subjects A group of 12 subjects (age range 20–32; 7 females, 5 males) took part in the experiment. All were right-handed and had normal or corrected-to-normal vision.

Stimuli and task In contrast to Experiment 1, informative cues predicted the relevant location of probe stimuli that had been present in the array with only 80% validity. Valid cues pointed to the correct location of the probe in the previous array, whereas invalid cues pointed to an incorrect

location that had previously been occupied by another stimulus in the array. There were three types of cues: valid, invalid, and neutral. To ensure sufficient trials in the invalid-cue condition, the number of SOAs between the array and cue presentations was reduced to three—150, 1,200, and 9,600 ms. These still spanned both IM and VSTM retention intervals.

There were 432 trials (360 informative and 72 neutral; i.e., 144 total at each SOA). Of the informative trials, 144 were valid (probe was in the array at the cued location), 36 were invalid (probe was in the array at an uncued location), and 180 included a probe stimulus that was absent from the array. Of the neutral trials, 36 included a probe that was in the array, and 36 included a probe that was not in the array. There were 12 blocks of trials in the experiment, plus an additional practice block at the beginning.

Results

Accuracy data were converted into d' scores, as in Experiment 1. These scores and the RT data were each analysed using an ANOVA testing for the behavioural benefits and costs of valid and invalid cues relative to neutral cues at the different SOAs. Since our validity manipulation could only work for target-present trials, only these were included in our analyses (a target-absent trial was necessarily neither validly nor invalidly cued).

Accuracy The d' scores were compared using a two-way ANOVA (these scores can be seen in Fig. 3). There was a main effect of validity [$F(2, 22) = 4.647, p = .021$]. The d' scores were higher with valid (1.97 ± 0.22) than with invalid (1.56 ± 0.20) [$F(1, 11) = 5.984, p = .032$] or neutral (1.49 ± 0.32) retro-cues [$F(1, 11) = 9.856, p = .009$]. The d' scores for invalid and neutral trials did not differ significantly [$F(1, 11) = 0.135, p = .720$]. There was a tendency towards an effect of SOA [$F(1.37, 15.12) = 2.934, p = .098$]; the pattern showed a linear trend of decreasing d' scores with increasing SOA, but this did not reach significance [$F(1, 11) = 3.380, p = .093$].

The interaction between validity and SOA also did not reach significance [$F(4, 44) = 2.150, p = .090$], although cue validity did interact significantly with the quadratic effect of SOA [$F(1, 11) = 9.187, p = .011$]. To isolate the benefits and costs of valid and invalid cues relative to neutral retro-cues, we produced difference scores and compared these across the various SOAs. The relative benefit of valid retro-cues was significantly modulated by SOA [$F(2, 22) = 5.158, p = .015$]. Follow-up t tests revealed that the benefit at SOA 1,200 ms was greater than the benefit at SOA 9,600 ms [$t(11) = 4.216, p = .001$], but not

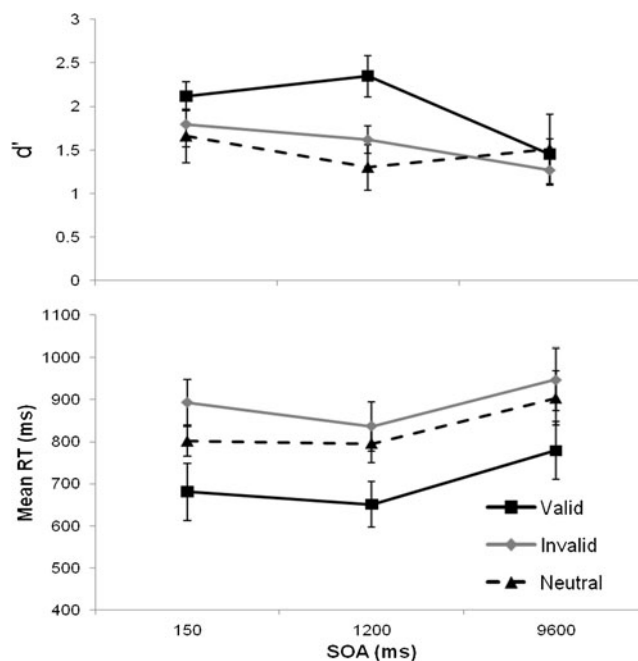


Fig. 3 Results from Experiment 2. The upper panel shows d' scores for valid, neutral, and invalid trials across the SOAs. The lower panel shows RTs for the same comparison. In both cases, the error bars show the standard errors of the means

significantly greater than the benefit at 150 ms [$t(11) = 1.505, p = .161$]. There was no significant difference between the benefits at SOAs of 150 and 9,600 ms [$t(11) = 1.411, p = .186$].

Reaction times Only RTs taken from correct trials were used in this analysis. The RT data were submitted to a repeated measures ANOVA, as in the d' analysis. The ANOVA revealed a main effect of cue type [$F(2, 22) = 21.914, p < .001$]. Post-hoc contrasts revealed that all of the cue conditions differed significantly from one another. RTs were faster in valid trials (703 ± 70 ms) than in neutral trials (834 ± 55 ms) [$F(1, 11) = 18.1, p = .001$] and invalid trials (893 ± 68 ms) [$F(1, 11) = 32.3, p < .001$]. RTs were slower in invalid trials than in neutral trials [$F(1, 11) = 6.534, p = .030$]. The pattern demonstrated both benefits of valid spatial cues and costs of invalid spatial cues as compared to neutral trials throughout the intervals tested. There was also a main effect of SOA [$F(2, 22) = 6.534, p = .006$]. Post-hoc contrasts indicated slower performance at the 9,600-ms SOA (877 ± 70 ms) than at the 1,200-ms SOA (761 ± 51 ms) [$F(1, 11) = 23.3, p = .001$], and marginally slower performance than at the 150-ms SOA (792 ± 58 ms) [$F(1, 11) = 4.3, p = .06$]. Performance at the two shorter SOAs (150 and 1,200 ms) differed, though this was only marginally significantly [$F(1, 11) = 3.6, p = .08$]. There was no significant interaction between cue and SOA [$F(4, 44) = 0.353, p = .840$].

Discussion

Whilst some previous studies have incorporated invalid trials (Astle et al., *in press*; Griffin & Nobre, 2003; Matsukura et al., 2007), to our knowledge this is the first to explore the extent to which these costs are modulated by cue SOA. Subjects were faster to find probe-matching coloured shapes in memory when they had been validly cued, relative to when a neutral cue was used. Subjects were also slower to find those shapes when they had been invalidly cued, also relative to a neutral-cue baseline. This was the case across all SOAs. However, the d' scores tell a subtly different story: Manipulating both SOA and cue validity appears to have had some unintended consequences. Because the array size was always within “capacity” (≤ 4 items), subjects could strategise their performance; by introducing invalid cues, one possibility was that subjects would attempt to maintain all items in case the cue transpired to be invalid. The d' scores would seem to support this; there was no d' cost of invalid relative to neutral cues, implying that subjects were not using the retro-cue to discard unwanted items. If subjects had already maintained the contents of VSTM for 9,600 ms, it would be potentially wasteful for them to select a single item for the last 500–1,000 ms when there was a possibility that they would select the wrong item—thereby reducing the effect of the retro-cue at the longer SOA. Indeed, not only was there no invalidity cost at SOA 9,600 ms, there was no validity benefit either—in marked contrast to Experiment 1.

The lack of accuracy costs on invalid trials relative to neutral trials in Experiment 2 would seem to contrast with the results of Matsukura et al. (2007), who demonstrated that invalid cues had a detrimental effect on accuracy measures, with both a double-cue and a single-cue paradigm. One possibility is that subjects use the cue in a strategic way, only discarding items if it is essential, and certainly not doing so if they have already maintained all of the array items for >9 s. That is, subjects might not always use a retro-cue to “protect” the cued VSTM item at the expense of the other items.

To summarise briefly, the RT data suggest that subjects used the retro-cue to direct their memory search to the most likely probe location, thereby speeding RTs on valid trials and slowing RTs on invalid trials. However, there was relatively little effect of cue validity on the accuracy data, and certainly there was no accuracy cost of invalid cueing. This implies that subjects did not use the retro-cue to discard uncued items, and thus to reduce the load (as in Matsukura et al., 2007), but rather to guide their memory search. Given that the array size was always within VSTM capacity (Cowan, 2001), when the cue was invalid on some trials, it was safest for subjects to encode *all* items into VSTM—thereby reducing the relative benefit of a valid

retro-cue, and eliminating any cost of invalid retro-cueing. Were this the case, we might expect the relative benefits and costs of retro-cueing to change as a function of load. When the array size exceeds capacity, subjects may have no choice but to use the cue to discard uncued items. The following experiment tested this possibility.

Experiment 3: Orienting attention to locations within large arrays in visual short-term memory

A traditionally held view is that VSTM, as assessed using a change detection paradigm, has a capacity of around four items (Cowan, 2001; Luck & Vogel, 1997; Vogel & Machizawa, 2004). To arrive at this number, the size of the memory array is varied, the proportion of nonchanges incorrectly detected (“false alarms”) is subtracted from the proportion of changes detected (“hits”), and the result is multiplied by the array size to produce an estimate of the number of items stored per array size, or K (Cowan, 2001). At around four items, subjects’ K typically asymptotes, implying that even though the arrays are getting progressively larger, the extra items are not being stored in VSTM. In some cases, the K estimate starts to decrease at around four items, implying that not only are the extra items not being stored, but they are actively interfering with the storage (or retrieval) of the other items.

In one previous study using spatial retro-cues, the authors argued that the true capacity of VSTM could be higher (even above 10 items) prior to the allocation of spatial attention, but that this fragile representation is overwritten at probe onset (Sligte et al., 2008). When a retro-cue is used, subjects are able to insulate the cued item from this overwriting (see also Makovski et al., 2008). Accordingly, Sligte et al. (2008) demonstrated that when the retro-cue is always valid, it could boost VSTM capacity to around 15 items. In short, they suggested that capacity is much higher than 4 items, provided that you can insulate those items from overwriting with spatial attention (as directed by a retro-cue). However, it is likely that the effect of the spatial retro-cues will be modulated by load. When load is low, or at least “within capacity,” subjects may use the retro-cue to prioritise memory search rather than discard items—as in Experiment 2. However, when load is high, or at least “exceeds capacity,” subjects may use the cue to discard uncued items from VSTM (as in Matsukura et al., 2007), thereby reducing the memory load to a single item. Experiment 3 explores this possibility.

In Experiment 3, we sought to explore two things: (1) We wanted to confirm previous findings that VSTM capacity is considerably higher, as measured using “supra-capacity” arrays, following valid cues (Sligte et al., 2008). (2) We sought to extend these findings by comparing both the

benefits and costs of spatial retro-cueing across different loads, to see whether the function of the cue would change depending on the array size. Experiment 3 used a VSTM task more typical of those experiments explicitly focused on the limits of VSTM capacity, with subjects having to detect changes at a probed location, rather than detect the presence of a probe at any location (the reasons for this change are explained in the Materials and Method section of this experiment).

Materials and method

The main new manipulation was the introduction of large arrays containing eight items, which are thought to exceed VSTM capacity (Cowan, 2001; Luck & Vogel, 1997; Vogel & Machizawa, 2004).

Subjects A group of 10 subjects (age range 19–32 years; 8 females, 2 males) took part in the experiment as volunteers.

Stimuli and task The task is shown in Fig. 4. In this experiment, the stimulus array was composed of two, four, or eight differently coloured crosses (33.3% probability). As in previous experiments, each trial began with a fixation point (400–700 ms), followed by the array of crosses (100 ms). Then, after a random interval between 1,500 and 2,500 ms, a cue was presented on the screen (100 ms), and after 500–1,000 ms the probe stimulus was displayed (100 ms).

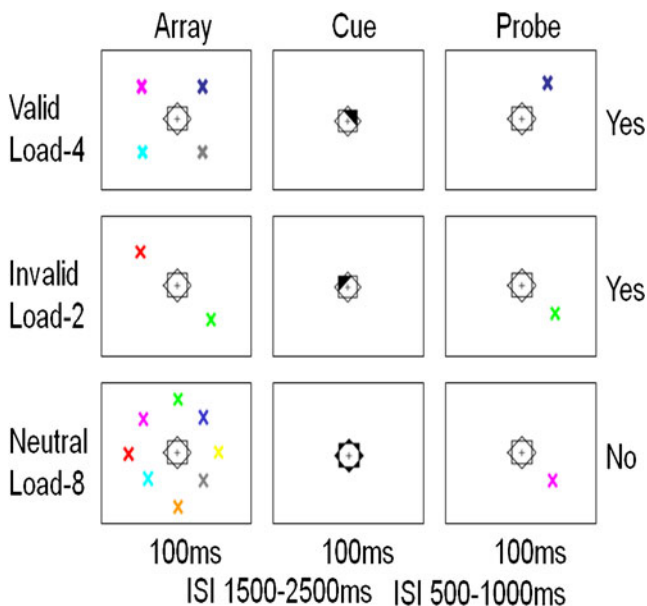


Fig. 4 A trial order schematic (Exp. 3) showing trials across three levels of load and three levels of cue validity. The upper two schematics are probe-present trials, and the lower schematic is a probe-absent trial

To make our paradigm more comparable to those typically used to estimate VSTM capacity, the probe appeared peripherally, at one of the previously occupied locations, and subjects were required to perform a change detection task. Subjects' task was therefore to decide whether the probe colour was the same as, or different from, that of the array shape previously presented at that location. They were instructed to decide ("yes" or "no") whether the colour of this probe matched the colour of the cross at that same location in the preceding array. In "no" trials, the probe item was always a colour that had been present in the array but that had appeared in a different location.

The change in paradigm from memory search (Exps. 1 and 2) to change detection carried several advantages for addressing the specific experimental questions. Any behavioural differences between valid, invalid, and neutral trials could not be affected by differential response criteria for different array locations, since decisions were probed at each location separately (Griffin & Nobre, 2003). In addition, the task ensured that subjects relied on memory about the item at a particular location, rather than on a general sense of familiarity or novelty about a particular colour. One additional, pragmatic reason was that, as load increased, the stimulus set size also increased in an unwieldy way. To maintain a memory search task, we would have needed 16 differently coloured items, making it difficult to differentiate the item colours because they would be so similar. With the paradigm used for Experiment 3, only eight items were needed, greatly reducing this problem. One potential difficulty with the change detection paradigm is that subjects may make intrusion errors—that is, whilst item information was maintained, locations might become confusable, especially when changes occurred at adjacent locations (see Chow, 1986, for a description).

There were three main trial types: valid, invalid, and neutral. Informative cues in this task predicted (80% validity) the location that would be probed. The cue consisted of two overlapping white squares whose corners each pointed to one of eight possible locations in the array. Valid and invalid informative cues (80% validity) consisted of one corner of the cue being filled in, pointing to a location that had been occupied in the preceding array. On valid trials, the cue correctly predicted the location of the array item that would be probed. On invalid trials, the cue incorrectly predicted the location of the array item that would be probed. Both colour-match and colour-nonmatch trials could be validly or invalidly cued. In neutral trials, all corners of the cue shape were filled in so that the cue gave no spatial information. For each array load (two, four, or eight items) and cue type (valid, invalid, or neutral), the probabilities of a correct colour match and a nonmatch were equal (50%).

There were 504 trials in total (336 valid, 84 invalid, 84 neutral). Of the valid trials, 56 were in each response and load condition (match–load 2, match–load 4, match–load 8, nonmatch–load 2, nonmatch–load 4, or nonmatch–load 8). In invalid and neutral trials, 14 were in each response and load condition (match–load 2, match–load 4, match–load 8, nonmatch–load 2, nonmatch–load 4, or nonmatch–load 8). There were 14 blocks of trials in the experiment, plus 1 additional practice block at the beginning. As in previous experiments, no feedback was given during the experiment.

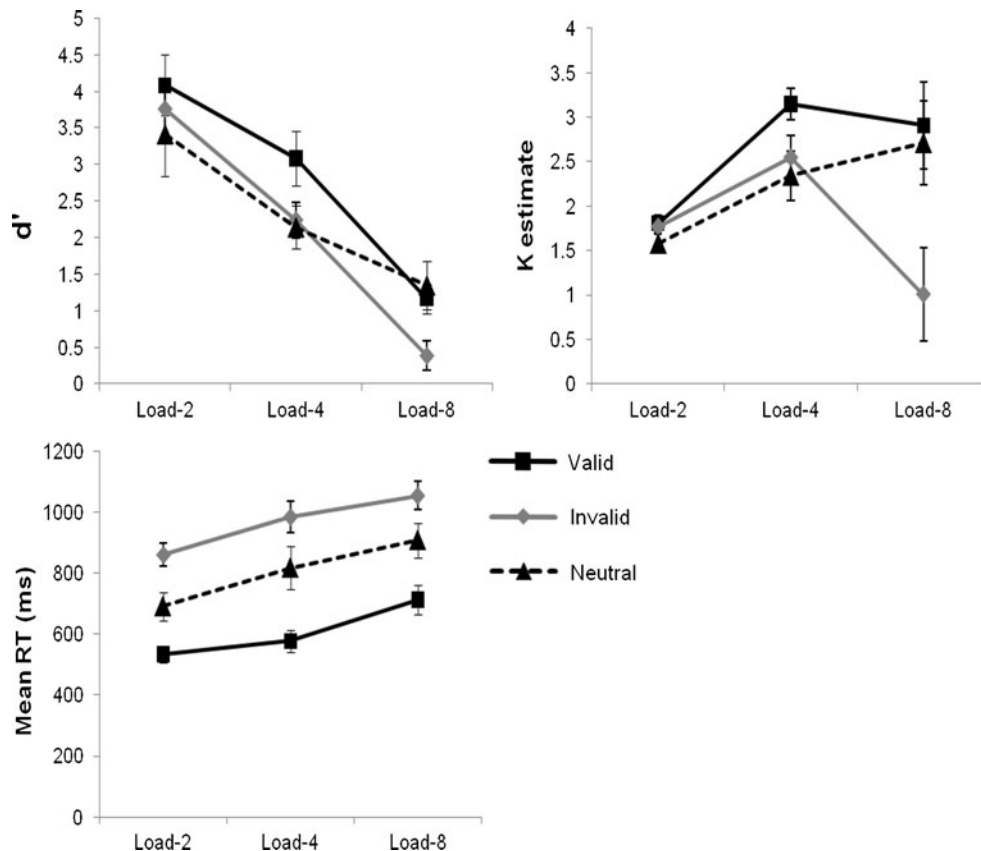
Results

The main aim of this experiment was to investigate the effect of retro-cues across the different VSTM loads. RT analyses focused only on trials on which the probed item had changed colour, though the *K* estimate and *d'* calculations also included no-change trials. Accuracy and RTs were analysed by repeated measures ANOVAs testing the factors of cue (valid, invalid, or neutral) and load (two, four, or eight items).

Accuracy We produced *d'* scores and submitted these to a two-way ANOVA, with the factors of load and validity (see the first panel of Fig. 5). There was a significant effect of load [$F(2, 18) = 37.799, p < .001$], with a significant linear

trend of decreasing *d'* scores with increasing load [$F(1, 9) = 78.767, p < .001$]. The *d'* value at load 2 (3.75 ± 0.45) was significantly higher than the values at load 4 (2.48 ± 0.33) [$F(1, 9) = 12.247, p = .007$] and load 8 (0.96 ± 0.28) [$F(1, 9) = 104.444, p < .001$]. The *d'* scores were also higher for load 4 than for load 8 [$F(1, 9) = 31.351, p < .001$]. However, there was also a main effect of validity on *d'* scores [$F(1.31, 11.75) = 4.503, p = .048$], with significantly higher scores on valid than on invalid trials [$F(1, 9) = 7.933, p = .020$], but no significant difference between neutral and invalid trials [$F(1, 9) = 0.363, p = .561$]. There was a significant decrease in *d'* scores from valid to neutral trials [$F(1, 9) = 13.287, p = .005$]. Importantly, there was also a significant interaction between validity and load [$F(4, 36) = 3.100, p = .027$]. No significant effect of cue type was found on load-2 trials [$F(2, 18) = 1.123, p = .347$], but there was on both load-4 [$F(2, 18) = 5.150, p = .017$] and load-8 [$F(2, 18) = 13.120, p < .001$] trials. At load 4, the effect stemmed from a significant benefit for valid retro-cueing, relative to either the invalid [$t(9) = 2.377, p = .041$] or the neutral [$t(9) = 2.739, p = .023$] condition. There was no significant difference between the invalid and neutral conditions [$t(9) = 0.381, p = .712$]. At load 8, the effect stemmed instead from a significant cost for invalid retro-cueing, relative to both the valid [$t(9) = 6.019, p < .001$] and neutral [$t(9) = 4.045, p =$

Fig. 5 Results from Experiment 3. The upper left panel shows *d'* across loads for the three levels of cue validity; the upper right panel shows *K* estimates (proportions of hits minus proportions of false alarms, multiplied by load) for the same comparison; the lower panel shows RTs for the same comparison. In all cases, the error bars show the standard errors of the means



.003] conditions. At load 8, there was no significant difference between the valid and neutral conditions [$t(9) = 0.832, p = .427$].

Capacity estimates We also used our accuracy data to produce K estimates. These were calculated by subtracting the proportion of false alarms from the proportion of correct hits for each load, and then multiplying this by the set size (Cowan, 2001). The peak in K across the various loads is usually taken as an estimate of that subject's capacity (Fukuda & Vogel, 2009; Vogel & Machizawa, 2004), but in this case we included load as a factor. These data can be seen in Fig. 5. The statistics are reported using the Greenhouse–Geisser correction for nonsphericity. There was a significant main effect of load [$F(1.25, 11.27) = 4.567, p = .049$], though this was to be expected, given that load was included in the calculation of K . More interestingly, we found a significant main effect of cue type [$F(1.38, 14.43) = 7.423, p = .012$], with K being higher on valid than on invalid trials [$F(1, 9) = 18.398, p < .001$] and also higher on valid than on neutral trials [$F(1, 9) = 5.519, p = .043$]. Invalid trials did not incur a significant cost relative to neutral trials [$F(1, 9) = 2.708, p = .134$]. Importantly, these two factors interacted significantly [$F(2.30, 20.69) = 7.371, p = .003$]. This was because we found no effect of cue type at load 2 [$F(1.32, 11.84) = 2.059, p = .177$] but did find effects at load 4 [$F(1.74, 15.66) = 5.845, p = .015$] and load 8 [$F(1.72, 15.51) = 7.961, p = .005$]. At load 4, valid retro-cues produced a higher K estimate than was present in either the neutral condition [$t(9) = 3.471, p = .007$] or the invalid retro-cue condition [$t(9) = 2.866, p = .019$]. At load 4, neutral and invalid cues did not differ significantly [$t(9) = 0.727, p = .486$]. At load 8, a significant effect of cue type arose because invalid retro-cues produced a reduced K estimate relative to both valid retro-cues [$t(9) = 3.775, p = .004$] and neutral cues [$t(9) = 2.829, p = .020$]. At load 8, neutral and valid retro-cues did not differ significantly [$t(9) = 0.276, p = .789$].

In summary, the costs and benefits of invalid and valid retro-cues were modulated by load. At load 4, subjects derived a relative benefit from the valid retro-cues but no cost from the invalid retro-cues. At load 8, valid retro-cues did not provide a benefit relative to the neutral retro-cues, but invalid retro-cues did elicit a significant cost. When the load was just within VSTM capacity (load 4), valid spatial retro-cues provided a benefit relative to both invalid and neutral cues. By contrast, when VSTM capacity was “exceeded,” we found a significant cost to invalid retro-cueing relative to both the valid and neutral conditions. This pattern of effects was apparent both in the d' and K scores.

Reaction times Only RTs from correct trials were used in this analysis. The pattern of RTs is shown in Fig. 5. The

ANOVA showed a main effect of cue [$F(2, 18) = 106.210, p < .001$]. Post-hoc contrasts revealed that responses were significantly faster on valid trials (608 ± 45 ms) than on neutral trials (806 ± 53 ms) [$F(1, 9) = 184.614, p < .001$] and invalid trials (967 ± 63 ms) [$F(1, 9) = 138.988, p < .001$]. Furthermore, RTs were significantly slower on invalid trials than on neutral trials [$F(1, 9) = 37.945, p < .001$]. Load also exerted a main effect [$F(2, 18) = 37.373, p < .001$], with a significant linear trend for increasing RTs with increasing load [$F(1, 9) = 73.119, p < .001$]. Post-hoc analyses showed that RTs slowed significantly for each load increment. Responses in load-2 trials (695 ± 56 ms) were significantly faster than those in load-4 trials (794 ± 76 ms) [$F(1, 9) = 27.414, p = .001$] and load-8 trials (895 ± 67 ms) [$F(1, 9) = 73.119, p < .001$]. Load-4 trials also had significantly faster RTs than did load-8 trials [$F(1, 9) = 14.463, p = .004$]. There was no significant interaction between validity and load [$F(4, 36) = 1.098, p = .373$].

Discussion

The results show that spatial orienting to items within VSTM representations modulated retrieval processes. As with cue SOA in the previous experiments, the relative RT benefits and costs of valid and invalid cueing were significant across all levels of load. However, in contrast with previous studies (e.g., Sligte et al., 2008), we only observed significant d' benefits when the array size was within capacity; when the array size was beyond capacity (eight items), valid cueing inferred no benefit relative to neutral trials. Conversely, only when capacity was exceeded did we observe a substantial *cost* of invalid cueing. This pattern of effects was present in both the d' and K measures.

Sligte et al. (2008) demonstrated that the capacity prior to attention being committed is much higher than the traditional four items. They compared performance on retro-cue and postcue conditions and demonstrated that the capacity of VSTM is at least double the traditional VSTM capacity, but that the apparent capacity drops to four items only on the presentation of the probe array. However, the results of Experiment 3 seem inconsistent with this view; valid retro-cues should have provided a benefit even when capacity was “exceeded,” whereas the benefit was limited to within-capacity loads. Not only were capacity estimates never boosted beyond four items, when the array size was eight items the valid retro-cue conferred no significant benefit, relative to neutral trials.

An additional contribution of Experiment 3 is that it explored the relative *costs* of this retro-cue-driven orienting. The pattern of results suggested that when load exceeds capacity, subjects use the cue to discard those uncued items

from VSTM, which has a catastrophic effect on accuracy when the cue is invalid; indeed, subjects performed worse than if they had attempted to maintain all of the array items (neutral trials). The results of Experiment 3 therefore seem to support and extend the findings of Matsukura et al. (2007): The use of the retro-cue differs, depending on whether VSTM capacity is exceeded; only when it is exceeded will subjects resort to using the retro-cue to discard items.

At load 2, there appears to be little effect of cue validity. This is unlikely to be because cueing benefits cannot be observed at such a low set size (see Nobre et al., 2008), but because subjects opt to use the cue only when the benefits of doing so outweigh the potential costs. If subjects can maintain the items in VSTM, there may be little incentive to use the cue, given that it might transpire to be invalid. Again, we suggest that subjects use the retro-cue strategically; if there were no potential cost to using the retro-cue, we would expect a cueing benefit, even at load 2 (as in Nobre et al., 2008).

Matsukura et al. (2007) observed the relative costs of a single invalid cue and of an initial invalid cue in the double-cue paradigm, at loads of both 4 and 6. This would seem to contrast with our interaction between VSTM load and cue validity. Why would subjects use the cue differently within and beyond VSTM capacity in our Experiment 3, and yet not do so in Matsukura et al.'s Experiment 3? We suspect that to get this pattern of costs across different loads, subjects need to experience all loads. Matsukura et al.'s comparison of load 4 and load 6 was between subjects, whereas our comparison of load 4 and load 8 was within subjects. Subjects may only demonstrate a differential strategy of cue use if they themselves experience the different memory loads. Furthermore, Matsukura et al. did not include any neutral trials, making it difficult to ascertain whether they were observing a validity benefit or an invalidity cost.

Experiment 4: Orienting attention within large arrays in IM and VSTM

Across the preceding three experiments, we explored two potential constraints on our ability to orient attention within mental representations. The first was the temporal constraint of whether the representation was held in IM or VSTM, and for how long the array had been held in VSTM. The second was the constraint of load. Experiment 3 showed a very interesting pattern of accuracy costs and benefits across loads, with relative benefits, but no costs, when cueing at capacity, and the reverse pattern when load exceeded capacity. Experiment 4 explored whether or not this relationship between load and retro-cueing benefits/

costs would interact with cue SOA, by investigating whether similar modulations across loads were observed within IM time spans.

Previous research demonstrated that the capacities of IM and VSTM differ greatly (Averbach & Coriell, 1961; Sperling, 1960). This was supported by a recent study showing that a valid retro-cue could boost VSTM capacity estimates to around 15 items, but that when operating within IM, K could reach around 20 items. When the stimuli leave an afterimage, the apparent capacity of IM could be even higher (Slight et al., 2008). With these findings in mind, our first aim was to replicate this effect and test whether capacity measures could be boosted most when the retro-cue operates on an IM, relative to a VSTM, representation. Secondly, we aimed to test whether the pattern of costs across loads seen in Experiment 3 would also occur when the retro-cue operated on an IM representation.

Materials and method

Experiment 4 used high-load arrays (eight items) as well as variation of the interval between the array and the cue, including retention intervals spanning IM and VSTM intervals. Unless stated otherwise, the materials and method were identical to those in Experiment 3.

Subjects A group of 10 subjects (age range 19–32 years; 8 females, 2 males) took part in the experiment as volunteers.

Stimuli and task The basic task is shown in Fig. 4. The main difference in this experiment relative to Experiment 3 was that the interval between the appearances of the array and the cue was either 150 ms (IM) or 1,500 ms (VSTM). In addition, only two array loads (four or eight items, 50% probabilities) were used. Cues could be valid, invalid, or neutral.

There were 480 trials in total (320 valid, 80 invalid, and 80 neutral). For valid cues, 40 trials were in each experimental cell (match–load 4–SOA 150 ms, nonmatch–load 4–SOA 1,500 ms, nonmatch–load 4–SOA 150 ms, nonmatch–load 4–SOA 1,500 ms, match–load 8–SOA 150 ms, nonmatch–load 8–SOA 1,500 ms, nonmatch–load 8–SOA 150 ms, nonmatch–load 8–SOA 1,500 ms). For invalid and neutral cues, there were 10 trials in each cell.

Results

As in Experiment 3, the RT analysis focused on the trials on which the probe item colour had changed (i.e., nonmatch trials), but the K estimate and d' calculations used both

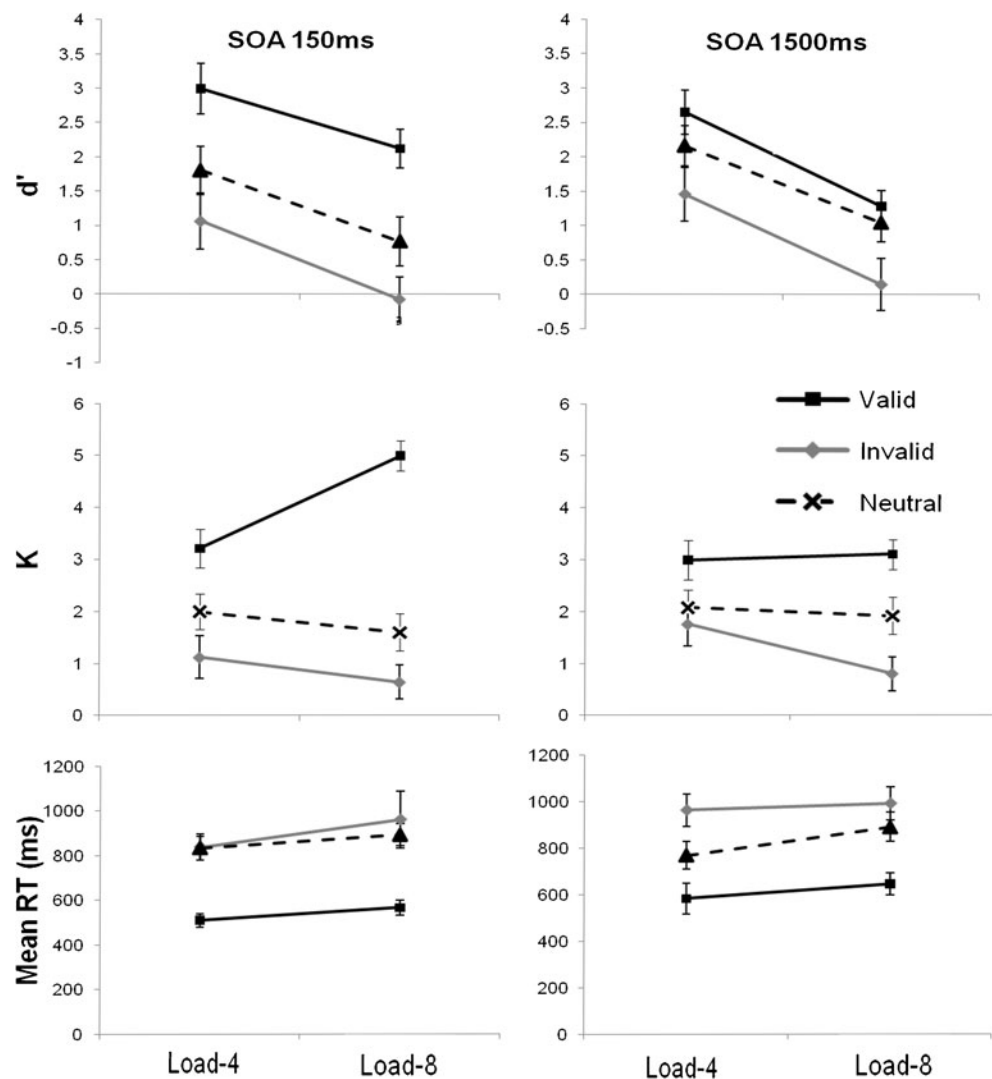
correct hits and false alarms. All of the effects were assessed by ANOVAs testing the factors of cue (valid, invalid, or neutral), load (four or eight items), and SOA (150 or 1,500 ms).

Accuracy The accuracy scores (see Fig. 6) were converted into d' scores. We entered these scores to a three-way ANOVA, with the within-subjects factors of SOA, load, and validity. There was a significant main effect of load, with scores being higher for load-4 (2.02 ± 0.40) than for load-8 (0.88 ± 0.38) trials [$F(1, 9) = 56.536, p < .001$]. There was a significant main effect of validity [$F(2, 18) = 57.518, p < .001$], with scores being higher for valid (2.26 ± 0.36) than for neutral (1.45 ± 0.41) [$F(1, 9) = 52.547, p < .001$] and for invalid (0.65 ± 0.37) [$F(1, 9) = 109.299, p < .001$] trials. The d' scores on invalid trials were also significantly lower than those on neutral trials [$F(1, 9) = 20.230, p < .001$]. The only significant interaction was between SOA and validity [$F(2, 18) = 7.274, p = .005$].

This resulted from a significant validity effect at SOA 150 ms [$F(2, 18) = 37.130, p < .001$] that stemmed from both a benefit of valid relative to neutral retro-cues [$F(1, 9) = 50.799, p < .001$] and a cost of invalid relative to neutral retro-cues [$F(1, 9) = 11.241, p = .008$]. There was also a significant difference between the valid and invalid conditions [$F(1, 9) = 48.797, p < .001$]. At SOA 1,500 ms, the significant effect of validity [$F(2, 18) = 29.264, p < .001$] was driven only by a small benefit of valid relative to neutral retro-cues [$F(1, 9) = 5.356, p = .046$], but a substantial cost of invalid retro-cueing relative to the neutral condition [$F(1, 9) = 17.716, p = .002$]. There was also a significant difference between the valid and invalid conditions [$F(1, 9) = 114.594, p < .001$].

Capacity measure We also included the K estimates across loads in a repeated measures ANOVA, with the factors of SOA, validity/cue type, and load (these data can be seen in Fig. 6). There was a significant main effect of cue

Fig. 6 Results from Experiment 4. The upper two panels show d' across the two levels of load, for iconic memory (left) and VSTM (right) SOAs, with three levels of cue validity; the middle two panels show the same comparison for K estimates (proportions of hits minus proportions of false alarms, multiplied by load); the lower two panels show the same comparison for mean RTs. In all cases, the error bars show the standard errors of the means



type [$F(2, 18) = 56.342, p < .001$], with K being higher on valid trials than on neutral [$F(1, 9) = 56.112, p < .001$] and invalid [$F(1, 9) = 66.699, p < .001$] trials. There was no main effect of load [$F(1, 9) = 0.005, p = .947$] or of SOA [$F(1, 9) = 0.959, p = .353$]; however, there were a number of two-way interactions: First, we found a significant interaction between SOA and cue type [$F(2, 18) = 6.480, p = .008$], with the cues having a greater influence on K at the IM interval [$F(2, 18) = 40.211, p < .001$] than at the VSTM interval [$F(2, 18) = 21.255, p < .001$]. At the IM delay/interval, the effect of retro-cues was mainly carried by a relative benefit for valid relative to the neutral [$F(1, 9) = 51.963, p < .001$] and invalid [$F(1, 9) = 48.535, p < .001$] conditions, and the cost of invalid retro-cues, relative to the neutral condition, was also significant [$F(1, 9) = 8.882, p = .015$]. At the VSTM delay, the effect of retro-cues was also carried by significant differences between valid retro-cues and both the neutral condition [$F(1, 9) = 10.920, p = .001$] and the invalid condition [$F(1, 9) = 28.248, p < .001$]. There was also a significant K cost for invalid retro-cues relative to the neutral baseline [$F(1, 9) = 8.428, p = .018$]. There was a significant interaction was between load and cue type [$F(2, 18) = 6.353, p = .008$], because load had the greatest effect on valid trials [$t(9) = -2.362, p = .042$], with no significant effect of load on neutral [$t(9) = 0.839, p = .423$] or invalid [$t(9) = 1.902, p = .090$] trials. The two-way interaction between SOA and load approached significance [$F(1, 9) = 4.473, p = .064$], because although load did not have a significant effect on K estimates at either SOA, it had a positive effect on K at the IM interval [$t(9) = -1.419, p = .190$] and, as we saw in the previous experiment, a negative effect at the VSTM interval [$t(9) = 1.663, p = .131$]. There was no three-way interaction between SOA, cue type, and load [$F(2, 18) = 1.873, p = .182$].

In summary, retro-cueing modulated accuracy at both the IM and VSTM intervals. As in Sligte et al. (2008), the retro-cues had their greatest effect at the IM delay, although our capacity estimates were not boosted to nearly the same extent. This modulation was driven primarily by a relative benefit from valid spatial retro-cueing, especially when the array size exceeded VSTM capacity. The cue type also exerted a significant effect at the VSTM delay.

Reaction times Only RTs from correct trials were included in this analysis. The RT results are also shown in Fig. 6. The ANOVA showed a main effect of cue [$F(2, 18) = 44.656, p < .001$]. Post-hoc comparisons showed that RTs were significantly faster in valid (576 ± 55 ms) than in neutral (846 ± 72 ms) [$F(1, 9) = 48.549, p < .001$] or invalid (938 ± 102 ms) [$F(1, 9) = 116.697, p < .001$] trials, and that RTs were significantly slower in invalid than in neutral trials [$F(1, 9) = 5.970, p = .037$]. Overall, RTs were

faster in load-4 trials (749 ± 74 ms) than in load-8 trials (824 ± 86 ms) [$F(1, 9) = 12.968, p = .006$]. The main effect of SOA was not significant, and there were no significant interactions.

Discussion

The pattern of results at load 8 in the VSTM condition replicated the pattern of results in the previous experiment: Valid retro-cues appear to confer little or no advantage relative to neutral retro-cues, but invalid retro-cues confer a substantial cost; this can be seen most obviously in our d' estimates. However, the effect at the IM SOA was somewhat different: Valid retro-cues substantially boosted performance at load 8, as well as invalid cues eliciting a significant cost. As was the case in Experiment 3, when operating on VSTM representations, valid cues enabled subjects to achieve a capacity estimate of a little over three and a half items, whereas when those valid retro-cues operated on IM representations, subjects were able to achieve a mean capacity estimate of over five items.

Sperling (1960) and others have suggested that the capacity of IM is greater than that of VSTM. Accordingly, we might expect subjects to be able to make better use of a valid retro-cue when they can use it to operate on the larger number of items held in IM than when operating on the already restricted items held in VSTM. As was mentioned in Experiment 1, the difference between retro-cueing at VSTM and IM intervals is most apparent when VSTM capacity is exceeded. If valid retro-cues enable subjects to operate on fragile item traces (Averbach & Coriell, 1961; Sligte et al., 2008; Sperling, 1960), the use of such a cue might yield the greatest benefit when it can operate with IM rather than VSTM. Presumably, these fragile supra-capacity traces are highly prone to decay or interference (Averbach & Coriell, 1961; Sperling, 1960), meaning that they would be more available for selection by early relative to late spatial cues. Using invalid cues in Experiment 4 also enabled us to test whether the costs are equivalent across these two delays. They appear to be similar: The d' scores dropped close to chance when eight items were in the array and the cue was invalid, implying that across both VSTM and IM, when the size of the array exceeds VSTM capacity, subjects use the retro-cue to discard uncued items, making them unavailable for reinspection at probe onset. Of course, the reason for this drop could differ across these two intervals: At the IM SOA, the uncued items might be lost to decay (Sperling, 1960); at the VSTM SOA, the uncued items might be lost intentionally, because capacity is exceeded (as in Exp. 3 and Matsukura et al., 2007).

General discussion

These experiments provide robust and compelling evidence that spatial attention can be oriented within the domain of mental representations to enhance and modulate memory retrieval. The four experiments focused on two constraints on mental representations. The first of these was a temporal constraint. We compared attention orienting within IM and VSTM durations (Exps. 1, 2, and 4) and extended the VSTM duration to 9,600 ms to explore the effect of VSTM decay on retrospective attention orienting (Exps. 1 and 2). The second constraint was that of load. We explored the orienting of attention to locations within either VSTM (Exps. 3 and 4) or IM (Exps. 1, 2, and 4) representations when they were within either the supposed capacity limit of VSTM (loads of 2–4 items) or beyond the capacity limit of VSTM (load 8). The final experiment brought these two constraints together, exploring any effects of load on the orienting of attention within IM and VSTM representations. Our aim was to test whether these factors might influence the mechanisms through which retro-cues improve performance.

There were a number of important results: (1) When the informative spatial retro-cue was 100% valid and the set size was within VSTM capacity, it enhanced retrieval over a wide range of SOAs, biasing representations in both IM and VSTM, even at cue SOAs of ~10 s (Exp. 1). (2) When the number of to-be-remembered items did not exceed VSTM capacity, subjects did not use the retro-cue to discard uncued items; even when cued to the wrong item, subjects' retrieval was slower but no less accurate (Exps. 2, 3, and 4). (3) When the number of items exceeded VSTM capacity (eight items), subjects used the retro-cue to discard uncued items; when cued to the wrong item, accuracy measures dropped to little better than chance levels (Exps. 3 and 4). (4) This pattern of benefits and costs within and beyond capacity did not hold when using the retro-cue to operate on an IM representation (Exp. 4). When operating within IM, valid retro-cues conferred accuracy benefits both within and beyond "capacity"; indeed, the greatest benefit was seen with supra-VSTM-capacity arrays.

It has been shown that a valid retro-cue can provide a K benefit relative to a postcue, even in arrays of 32 items (Sligte et al., 2008). Given this finding, it is surprising that whilst subjects were *faster* to retrieve the validly cued item, we did not observe a K or d' benefit in arrays of eight items (Exp. 3). Our data therefore do not support the existence of a pool of items, greater than the capacity of VSTM, that can be maintained using a retro-cue. As we mentioned earlier, we suspect that one reason for this is that subjects will not rely on the cue to the same extent when they suspect that it might be invalid. Thus, whilst our designs enabled us to explore the relative costs of retro-cueing to those items

uncued, these designs might also preclude us from observing the large benefits seen by Sligte et al. (2008). A possible secondary cause of this apparent difference is the degree of prior training that subjects were given. Because task instructions were relatively simple, we gave subjects around 50 trials practise; Sligte et al. (2008) gave their subjects substantially more training (around 3 h). Highly trained subjects who rely entirely on the retro-cue might be required in order to see these massive increases in K estimates by retro-cues. However, a more recent study from the same group fits well with our findings: Sligte et al. (2010) used 100% valid retro-cues (as in Exp. 1), delivered at different SOAs, to probe the capacity of different short-term stores. As in Experiments 3 and 4, the array size was eight items. Delivering the cue within an IM period (10 ms postarray) yielded a K estimate of a little over six items, which dropped to a little over four items when the cue was delivered within a VSTM period (1,000 ms postarray), and dropped farther still, to a little over two items, in a non-retro-cued version (in this control condition, the cue was delivered after the change had occurred). These K estimates are similar to those that we observed in Experiment 4: When operating on eight items in IM (cue SOA = 150 ms), retro-cues yielded a K estimate of a little over five items, which dropped to a little over three items when the cue was delivered in a VSTM period (1,500 ms postarray), and dropped farther still to around two items when no retro-cue was used (the neutral conditions). Our K estimates are a little lower than those observed by Sligte et al. (2010), but this is perhaps to be expected: Sligte et al.'s (2010) cues were delivered slightly earlier, and they used real-life objects rather than easily confusable coloured crosses.

Our results extend previous findings in two respects: Firstly, we used neutral cues in order to establish the *relative benefits* of these different retro-cue SOAs. The retro-cue is most effective at boosting capacity estimates when it operates on IM, consistent with the view that these items are most easily lost without such cues (Averbach & Coriel, 1961; Sperling, 1960). The retro-cue provides little *benefit*, in terms of boosting capacity, when it operates on a VSTM representation. Secondly, our use of invalid cues demonstrates that retro-cues operating on "supra-capacity" arrays are effective because, at least partially, they enable subjects to discard uncued items, thereby reducing the load. This is true across all SOAs.

As in previous studies, it is difficult to ascertain how items are lost from VSTM, and we certainly make no claim here as to how this arises. However, to our knowledge, for the first time, the present study demonstrates that the retro-cue is not always used in the same way. When VSTM capacity is not exceeded, subjects use the retro-cue to prioritise where they initiate their memory search, with

retrieval being slowed but not less accurate following an invalid retro-cue (see also Nobre et al., 2008). It is only when VSTM capacity is exceeded that subjects use the cue to reduce the number of stored items. Although it remains to be seen how subjects do this and whether the removal of items is achieved through the same means in IM and VSTM. One possibility is that uncued items in IM are not insulated from decay, which, given the fragile nature of these representations, will mean that they are lost by probe onset (Sperling, 1960); by contrast, when these items are stored in VSTM, subjects may actively remove the previously applied insulation, or actively suppress the item representation.

A final important issue to consider, which may differentiate this study from some previous examples in the literature, is whether subjects emphasise speed or accuracy. In our experiments, we did not emphasise one over the other, since we were looking for cueing effects in both speed and accuracy. Other example studies (e.g., Matsukura et al., 2007) have explicitly emphasised accuracy over speed to subjects. This difference may well affect the mechanisms at play. We are confident that our effects of load, validity, and SOA in d' or K do not stem simply from differential speed–accuracy trade-offs across these different factors; in no case did the RT effects reflect the inverse of the accuracy effects.

Author Note D.E.A. was supported by a Postdoctoral Fellowship from the British Academy and by project grants from the John Fell Fund (Oxford University Press) and the Royal Society. A.C.N. was supported by a 21st Century Research Award from the James S. McDonnell Foundation (Bridging Brain, Mind and Behaviour) and by a project grant from the Wellcome Trust. We thank Morag Brothwell for her help in collecting preliminary data that led to these experiments. We also thank Bo-Cheng Kuo, Ilja Sligte, and three anonymous reviewers for their comments on earlier drafts.

References

- Astle, D.E., Nobre, A. C., & Scerif, G. (in press). Attentional control constrains visual short-term memory: Insights from developmental and individual differences. *Quarterly Journal of Experimental Psychology*. doi:10.1080/17470218.2010.492622
- Astle, D. E., Scerif, G., Kuo, B.-C., & Nobre, A. C. (2009). Spatial selection of features within perceived and remembered objects. *Frontiers in Human Neuroscience*, 3, 6. doi:10.3389/neuro.09.006.2009
- Averbach, E., & Coriell, A. S. (1961). Short-term memory in vision. *Bell Systems Technical Journal*, 40, 309–328.
- Bays, P. M., Catalao, R. F., & Husain, M. (2009). The precision of visual working memory is set by allocation of a shared resource. *Journal of Vision*, 7(10), 1–11.
- Cowan, N. (2001). The magical number 4 in short-term memory: A reconsideration of mental storage capacity. *The Behavioral and Brain Sciences*, 24, 87–114. doi:10.1017/S0140525X01003922
- Chow, S. L. (1986). Iconic memory, location information, and partial report. *Journal of Experimental Psychology: Human Perception and Performance*, 12, 455–465. doi:10.1037/0096-1523.12.4.455
- Dick, A. O. (1974). Iconic memory and its relation to perceptual processing and other memory mechanisms. *Perception & Psychophysics*, 16, 575–596.
- Fukuda, K., & Vogel, E. K. (2009). Human variation in overriding attentional capture. *Journal of Neuroscience*, 29, 8726–33.
- 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
- 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
- Lee, D., & Chun, M. M. (2001). What are the units of visual short-term memory, objects or spatial locations? *Perception & Psychophysics*, 63, 253–257. doi:10.3758/BF03194466
- Lepsien, J., Griffin, I. C., Devlin, J. T., & Nobre, A. C. (2005). Directing spatial attention in mental representations: Interactions between attentional orienting and working-memory load. *NeuroImage*, 26, 733–743. doi:10.1016/j.neuroimage.2005.02.026
- Lepsien, J., & Nobre, A. C. (2006). Cognitive control of attention in the human brain: Insights from orienting attention to mental representations. *Brain Research*, 1105, 20–31. doi:10.1016/j.brainres.2006.03.033
- 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., Shim, W. M., & Jiang, Y. V. (2006). Interference from filled delays on visual change detection. *Journal of Vision*, 11(12), 1459–1470. doi:10.1167/6.12.11
- 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
- Neisser, U. (1967). *Cognitive psychology*. New York: Appleton-Century-Crofts.
- 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
- O'Regan, J. K., Rensink, R. A., & Clark, J. J. (1999). Change-blindness as a result of “mudsplashes”. *Nature*, 398, 34. doi:10.1038/17953
- Posner, M. I. (1980). Orienting of attention. *Quarterly Journal of Experimental Psychology*, 32, 3–25. doi:10.1080/0033558008248231
- Rensink, R. A., O'Regan, J. K., & Clark, J. J. (1997). To see or not to see: The need for attention to perceive changes in scenes. *Psychological Science*, 8, 368–373. doi:10.1111/j.1467-9280.1997.tb00427.x
- Sligte, I. G., Scholte, H. S., & Lamme, V. A. (2008). Are there multiple visual short-term memory stores? *PLoS One*, 3, e1699.
- Sligte, I. G., Scholte, H. S., & Lamme, V. A. (2009). V4 activity predicts the strength of visual short-term memory representations. *Journal of Neuroscience*, 29, 7432–7438.
- Sligte, I. G., Vandenbrouke, A. R. E., Scholte, H. S., & Lamme, V. A. F. (2010). Detailed sensory memory, sloppy working memory. *Frontiers in Psychology*, 1, 175.

- Sperling, G. (1960). The information available in brief visual presentation. *Psychological Monographs*, 74(11, Whole No. 498), 1–29.
- Sternberg, S. (1966). High-speed scanning in human memory. *Science*, 153, 652–654. doi:[10.1126/science.153.3736.652](https://doi.org/10.1126/science.153.3736.652)
- Vogel, E. K., & Machizawa, M. G. (2004). Neural activity predicts individual differences in visual working memory capacity. *Nature*, 428, 748–751.
- Vogel, E. K., Woodman, G. F., & Luck, S. J. (2001). Storage of features, conjunctions, and objects in visual working memory. *Journal of Experimental Psychology. Human Perception and Performance*, 27, 92–114. doi:[10.1037/0096-1523.27.1.92](https://doi.org/10.1037/0096-1523.27.1.92)
- Zhang, W., & Luck, S. J. (2008). Discrete fixed-resolution representations in visual working memory. *Nature*, 453, 233–235. doi:[10.1038/nature06860](https://doi.org/10.1038/nature06860)