Selection of multiple cued items is possible during visual short-term memory maintenance

Michi Matsukura · Shaun P. Vecera

Abstract Recent neuroimaging studies suggest that maintenance of a selected object feature held in visual short-term/ working memory (VSTM/VWM) is supported by the same neural mechanisms that encode the sensory information. If VSTM operates by retaining "reasonable copies" of scenes constructed during sensory processing (Serences, Ester, Vogel, & Awh, 2009, p. 207, the *sensory recruitment hypothesis*), then attention should be able to select multiple items represented in VSTM as long as the number of these attended items does not exceed the typical VSTM capacity. It is well known that attention can select at least two noncontiguous locations at the same time during sensory processing. However, empirical reports from the studies that examined this possibility are inconsistent. In the present study, we demonstrate that (1) attention can indeed select more than a single item during VSTM maintenance when observers are asked to recognize a set of items in the manner that these items were originally attended, and (2) attention can select multiple cued items regardless of whether these items are perceptually organized into a single group (contiguous locations) or not (noncontiguous locations). The results also replicate and extend the recent finding that selective attention that operates during VSTM maintenance is sensitive to the observers' goal and motivation to use the cueing information.

Keywords Visual attention · Visual short-term/working memory · Retention-interval cueing · Sensory recruitment hypothesis · Multiple cues

M. Matsukura (✉) · S. P. Vecera
Department of Psychology, University of Iowa, 11 Seashore Hall E, Iowa City, IA 52242, USA
e-mail: michi-matsukura@uiowa.edu

Recent neuroimaging studies (e.g., Albers, Kok, Toni, Dijkerman, & de Lange, 2013; Christophel, Hebart, & Haynes, 2012; Ester, Anderson, Serences, & Awh, 2013; Harrison & Tong, 2009; Pratte & Tong, 2014; Serences, Ester, Vogel, & Awh, 2009) have consistently shown that early sensory areas, such as area V1 (primary visual cortex), are recruited to maintain a selected object feature in visual short-term memory (VSTM).[1] Similar to maintenance of the object feature, Munneke, Belopolsky and Theeuwes (2012) showed that those early sensory areas are engaged as observers' attention was shifted among multiple spatial locations represented in VSTM. These findings suggest that memory representations in V1 are supported by the same neural mechanisms that encode the sensory information (i.e., the *sensory recruitment hypothesis*, e.g., Awh & Jonides, 2001). If VSTM operates "by retaining reasonable copies of scenes" constructed during sensory processing (Serences et al., 2009, p. 207), then attention should operate in similar manners across perceptual and memory representations *to a certain extent*, despite visual selections within each of these representations are functionally dissociable (e.g., Hollingworth & Maxcey-Richard, 2013). For instance, despite different modes of orienting (e.g., Jonides, 1981; Posner & Cohen, 1984), the visual system can utilize both exogenous and endogenous cues to direct attention to a particular location during sensory processing (see Yantis, 2000, for a review). Indeed, a recent study conducted by Matsukura, Cosman, Roper, Vatterott and Vecera (2014) demonstrated that, despite lack of luminance transients brought by the bona fide exogenous cue (such as the one observed during sensory processing), the exogenous (peripheral) cue can guide attention to a particular item's location represented in VSTM as efficiently as the endogenous (central) cue (but see Berryhill, Richmond, Shay, &

---

[1] We consider VSTM and visual working memory (VWM) as the same set of processes.
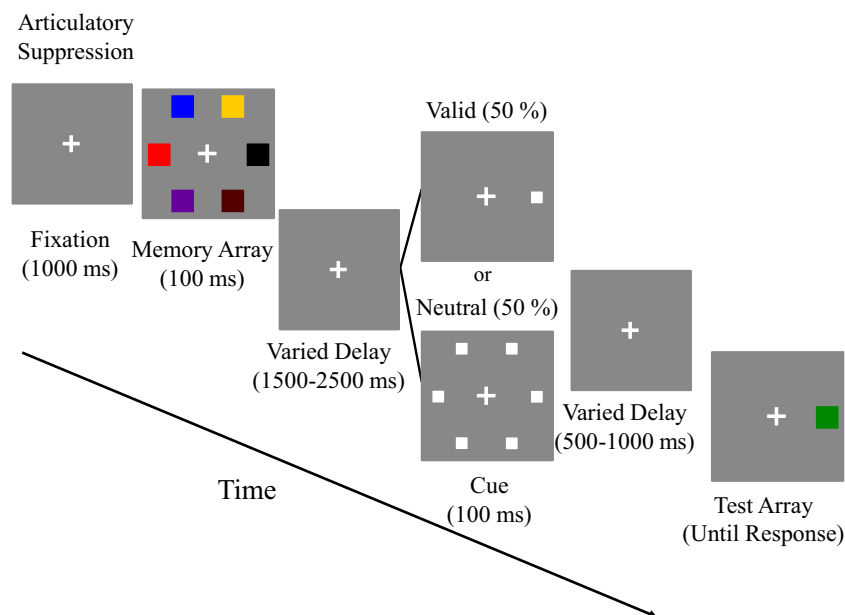
Fig. 1 Trial event sequence of a single-item retention-interval cueing paradigm with a peripheral dot cue (memory-array set size 6, different-color trial, Matsukura et al., 2014). This sequence also represents a cue-set-size-1 trial in all experiments of the present study (except Exp. 1B with the central arrow cue). Note that, for illustrative purpose, the stimuli are drawn much larger than they appeared in the actual computer display and some color values are adjusted (see Exp. 1A Method for actual color values). To view the figures in color, please see the online version of this article

Olson, 2012, for the account that the exogenous cue cannot guide attention during VSTM maintenance; thus, attention operates in a quintessentially different manner from "during sensory processing").[2]

In the present study, we asked another important question linked to the sensory recruitment hypothesis. That is, we investigated whether attention can select more than a single item during VSTM maintenance. If VSTM operates by keeping "reasonable copies of scenes evoked during sensory processing" (Serences et al., 2009, p. 207), then attention should be able to select multiple items represented in VSTM as long as the number of these attended items does not exceed the typical VSTM capacity of three to four items (Luck & Vogel, 1997; Vogel, Woodman, & Luck, 2001). It is well known that, during sensory processing, attention can select at least two non-contiguous locations at the same time (e.g., Alvarez, Gill, & Cavanagh, 2012; Anderson, Ester, Serences, & Awh, 2013; Awh & Pashler, 2000; Ester, Fukuda, May, Vogel, & Awh, 2014; Franconeri, Alvarez, & Enns, 2007; Hahn & Kramer, 1998; Kramer & Hahn, 1995). Yet, empirical reports from the studies that examined the multiple-item cueing effects during VSTM maintenance are inconsistent (Delvenne & Holt, 2012;

Makovski & Jiang, 2007; Matsukura, Luck, & Vecera, 2007; Williams & Woodman, 2012).

Since Griffin and Nobre (2003) reported that attention can select a visual item already stored in VSTM (i.e., even after an iconic image of to-be-remembered items faded away, but see Averbach & Coriell, 1961; Sperling, 1960), a number of studies replicated the retention-interval cueing effects (e.g., Astle, Summerfield, Griffin, & Nobre, 2012; Makovski & Jiang, 2007; Makovski, Sussman, & Jiang, 2008; Matsukura et al., 2007, 2014; Matsukura & Hollingworth, 2011; Murray, Nobre, Clark, Cravo, & Stokes, 2013; see Hollingworth & Hwang, 2013; Pertzov, Bays, Joseph, & Husain, 2013; Souza, Rerko, Lin, & Oberauer, 2014; Williams, Hong, Kang, Carlisle, & Woodman, 2013, for replications with nonbinary, continuous measures; see Landman, Spekreijse, & Lamme, 2003; Sligte, Scholte, & Lamme, 2008, for replications with all validly cued trials).

Both retention-interval cueing benefit and cost are typically measured by presenting a *single* attention-directing cue during the delay of a change-detection trial (similar to Fig. 1). A memory array that contains a set of colors is briefly presented. After a brief delay interval, a test array that contains a single test probe appears. During this delay period, a single cue appearing 500 ms or longer after the memory-array offset (i.e., after the iconic image of the memory array disappeared; Irwin & Yeomans, 1986; Sperling, 1960) either correctly or incorrectly predicts the to-be-tested item's location for a certain percentage of trials. Observers are asked to report whether a single test probe had the same or different color from the item

---

[2] As explicitly stated in Matsukura et al. (2014), the term "exogenous/peripheral" here is used to denote the condition that the attention-directing cue appears at the exact same perceptual location with the cued item represented in VSTM, not to suggest that such a cue unpredictably draws an observer's attention due to its abruptness or saliency (usually tested with the cue validity below 50 %).

presented at the corresponding location in the memory array. For the cueing benefit, the observers recognize validly cued items more accurately in comparison to neutrally cued items. For the cueing cost, the observers recognize invalidly cued items less accurately relative to neutrally cued items.

While the overwhelming consensus is that attention can select a single cued item already stored in VSTM,

the answer to the question, whether attention can select multiple cued items during VSTM maintenance or not, has been elusive. On the one hand, the results from the two studies suggest that attention can select multiple cued items during VSTM maintenance (Matsukura et al., 2007; Williams & Woodman, 2012). On the other hand, the results from a particular study indicate that
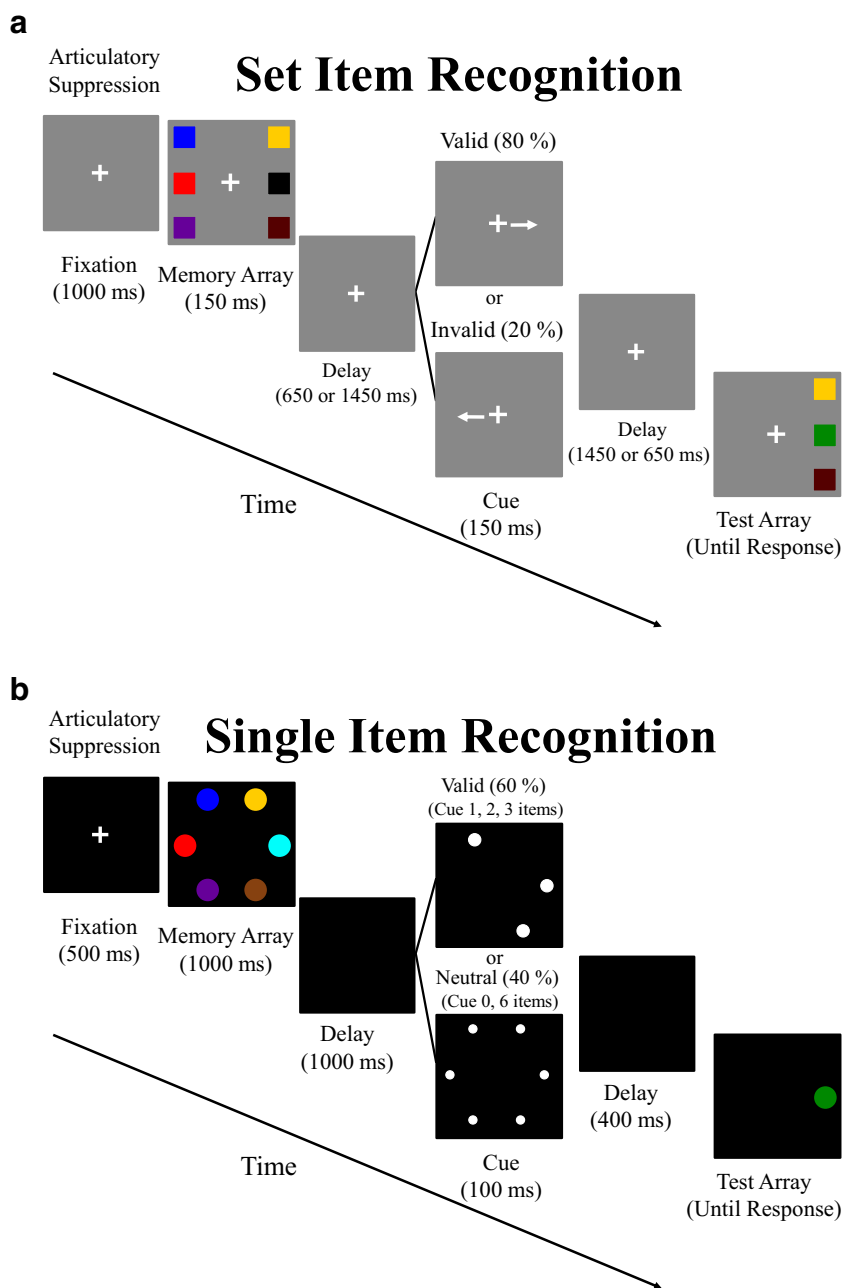

Fig. 2 a The set-item recognition task used in Matsukura et al. (2007, Exps. 3 and 4). Event sequence represents a single-cue trial of their double-cue procedure (memory-array set size 6, different-color trial). The observers were asked to recognize if there is any color change between the probed set of items in the test array and the cued set of items appeared at the corresponding locations in the memory array. **b** The single-item recognition task used in Makovski and Jiang (2007,

Exp. 2A, memory-array set size 6, different-color trial). The observers were asked to recognize if a single test probe had the same or different color from the item appeared at the corresponding location in the memory array, regardless of cue set size. Makovski and Jiang replicated the same pattern of results with the 800-ms delay between the cue offset and a single test probe onset

attention cannot select more than a single item once to-be-remembered items are stored in VSTM (Makovski & Jiang, 2007). What is so intriguing about this discrepancy is that, while significant multiple-item cueing benefits observed during VSTM maintenance are in line with the sensory recruitment account, the reported lack of the multiple-item cueing benefit is inconsistent with such an account.

Figure 2 contrasts the two retention-interval cueing tasks employed by the two different groups of studies mentioned above. In the context of examining how the retention-interval cueing effects were generated (i.e., protection or prioritization), Matsukura et al. (2007, Exps. 3 and 4, Exp. 5 for the control) directed the observers' attention to *a set of items* during VSTM maintenance. Figure 2a illustrates an event sequence of a single-cue trial from their procedure. At the time of the test, the observers were asked to report whether there was any color change between the probed set of items in the test array and the cued set of items appeared at the corresponding locations in the memory array (i.e., *set-item recognition*). If there was any change between these two sets of items, then this change was limited to a single item out of the three cued items. That is, the observers were asked to report if everything was the same between the probed and cued sets of items (*same-color* trial) or differed by one item (*different-color* trial). With this set-item recognition task, the robust retention-interval cueing effects were repeatedly observed. Williams and Woodman (2012) subsequently used a similar set-item recognition task and obtained a significant cueing benefit (the directed-remembering condition, compared against the baseline/no-cue condition in their Exp. 3).

Figure 2b illustrates the recognition task employed by Makovski and Jiang (2007, Exp. 2A). Investigating whether the observers could use multiple retention-interval cues, Makovski and Jiang presented zero, one, two, three, or six peripheral dot cues equiprobably during the retention interval and compared accuracy of each cue-set-size condition with accuracy of the no-cue (cue-set-size-0) condition. Because cue-set-size-6 trials are equivalent to neutrally cued trials (e.g., Matsukura et al., 2007, 2014), no significant difference was found between the cue-set-size-0 and -6 conditions. A significant cueing benefit was found in the cue-set-size-1 condition alone (compared against the cue-set-size-0 condition). Here, one striking difference from the set-item recognition task introduced earlier (Matsukura et al., 2007; Williams & Woodman, 2012) is that the observers' attention was directed to a set of three items (i.e., multiple items) during VSTM maintenance; however, at the time of the test, the observers were asked to report whether or not a single test probe shared the same color with "a single cued item" which had appeared at the corresponding location in the memory array, and this single cued item was "one of the multiple cued items" (i.e., *single-item recognition*).

Such a difference between the tasks raises the strong possibility that attention can actually select more than a single item during VSTM maintenance when the observers are asked to recognize a set of items in exactly the same way that these items were attended during VSTM maintenance. In other words, multiple retention-interval cues can be efficiently utilized when there is a reasonable contextual match between the cued set of items represented in VSTM and the probed set of items in the test array (i.e., in perception). However, there is an alternative account. So far in the retention-interval cueing literature, the studies that found the multiple-item cueing benefit directed the observers' attention to a single perceptually grouped set of items (Matsukura et al., 2007; Williams & Woodman, 2012). As illustrated in Fig. 2a, the cued set of items was grouped by proximity. In contrast, the study that failed to observe the multiple-item cueing benefit directed the observers' attention to the set of items placed on random noncontiguous locations in the memory array (Makovski & Jiang, 2007; see Fig. 2b). In Makovski and Jiang's case, because the cued items were not grouped into a single continuous region by proximity, attention could not be oriented *to a specific direction*. This stipulation raises the possibility that memory-level attention can select multiple cued items only under the circumstances that attention is directed to a single perceptually grouped set of items in VSTM. To distinguish these alternatives, the current series of experiments were designed in the way that enabled us to examine the effect of perceptual grouping on the multiple-item cueing benefit.

Upon linking the observers' ability to utilize multiple retention-interval cues to the sensory recruitment hypothesis, we should acknowledge two known properties of selective attention that operates during VSTM maintenance. Because the sensory recruitment account does not postulate that attention operates in exactly the same manner across memory and perceptual representations, the following characteristics of memory-level attention should be clearly acknowledged before going into specifics of the present study. First, in the present study, we assume that attention involved in the retention-interval cueing experiment operates within the limited capacity of three to four items (with canonical change-detection tasks developed by Luck & Vogel, 1997). This is consistent with the view that the retention-interval cueing benefit (and cost) is produced by preferential retention of the cued item relative to other uncued items within the limited VSTM capacity, without changing the quality of VSTM (certainly not for the direction of enhancing the VSTM quality—e.g., through protection and prioritization; Matsukura et al., 2007; Matsukura & Hollingworth, 2011; but see Sligte et al., 2008, for the account that use of the retention-interval cue allows the observers to access a high-capacity stage of VSTM). Second, we do not assume that selective attention that operates within VSTM representations (memory-level attention) and selective attention that operates during sensory processing (perception-level

attention) share a common *functional* mechanism (e.g., Hollingworth & Maxcey-Richard, 2013). As recently reviewed in Matsukura et al. (2014), although memory-level and perception-level selection mechanisms are functionally dissociable, a common set of attentional control mechanisms can still be used across many different types of attention tasks (e.g., Exp. 5 in Matsukura et al., 2007; Wojciulik & Kanwisher, 1999; see p. 1430 in Matsukura et al., 2007, for the original argument). Thus, the anticipated results that attention can select multiple cued items during VSTM maintenance, as observed during sensory processing, do not necessarily refute a functional dissociation between memory-level and perception-level selections.

Having acknowledged the two properties of memory-level attention, we will first examine whether or not attention can select multiple cued items during VSTM maintenance with the set-item recognition task (Exp. 1). It should be remembered that the theoretical argument that we make in the present study is that multiple retention-interval cues can be efficiently utilized as long as there is a reasonable contextual match between the cued representations in VSTM and the to-be-compared percepts in the test array. Such an argument would not be complete unless lack of the multiple-item cueing benefit is successfully replicated with the single-item recognition task. For this reason, the subsequent experiments will be dedicated to replicate Makovski and Jiang's (2007) results (Exps. 2–4).

To preview our results, consistent with the sensory recruitment hypothesis, selection of multiple cued items is possible during VSTM maintenance. As long as the observers were asked to recognize a set of items in the way that these items had originally been attended during VSTM maintenance, significant multiple-item cueing benefits were observed regardless of whether the cued items were perceptually grouped into a single continuous region or not. These results strongly suggest that lack of the multiple-item cueing benefit reported by Makovski and Jiang (2007) should not be interpreted as a quintessential feature of memory-level attention.

General method

All experiments in the present study were conducted with presenting six color items in the memory array. Besides the fact that this memory-array set size is the one used by Makovski and Jiang (2007), the set size of the to-be-remembered items should exceed the typical VSTM capacity for attention to be maximally utilized to select the cued item from VSTM (e.g., Matsukura et al., 2014; see the view of process-oriented attention in Luck & Vecera, 2002). Within this memory-array-set-size-6 setting, the following four specifications remained constant across the four experiments.

First, we used only valid and neutral cues as invalid cues do not readily distinguish the different effects of
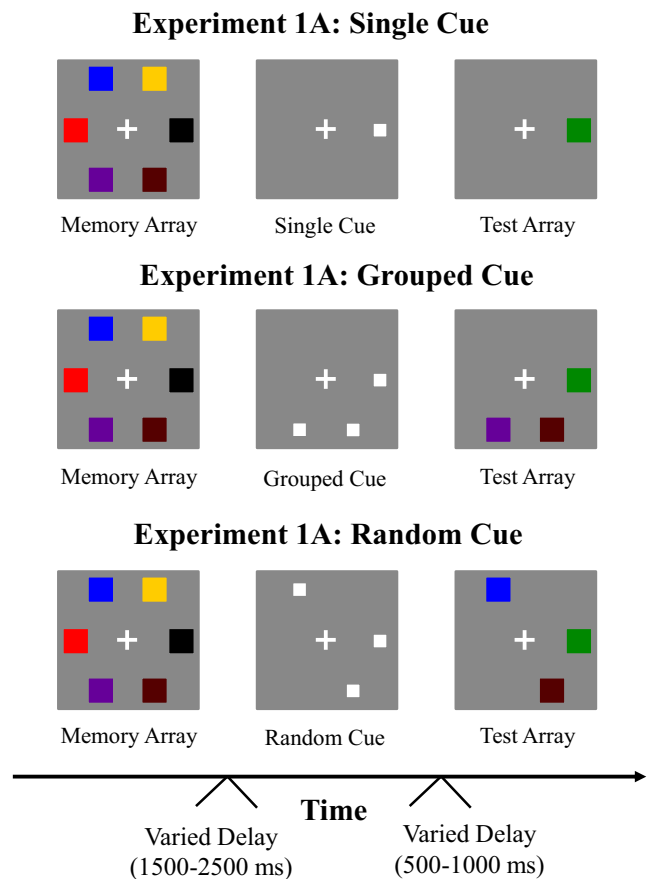

Experiment 1A: Single Cue

Memory Array   Single Cue   Test Array

Experiment 1A: Grouped Cue

Memory Array   Grouped Cue   Test Array

Experiment 1A: Random Cue

Memory Array   Random Cue   Test Array

Time

Varied Delay
(1500–2500 ms)

Varied Delay
(500–1000 ms)

Fig. 3 Event sequence of a validly cued trial for the single (top), grouped (middle) and random (bottom) cue conditions in Experiment 1A (the set-item recognition task)

attention. That is, we cannot distinguish whether the cueing cost is generated through the active use of the valid cue (as a researcher usually intends) or the observers intentionally forget the uncued items due to the demand characteristics of the experiment. We also made validly cued and neutrally cued trials equiprobable to increase statistical power (Fig. 1; see also Matsukura et al., 2007, 2014; Matsukura & Hollingworth, 2011, for this method). Neutrally cued trials were designed to diffuse attention across all six items represented in VSTM. The neutral cue was a set of small square dots that appeared at each of memory-array items' locations (except a set of central arrows used in Exp.1B), which is equivalent to cueing all six items.

Second, as illustrated in Fig. 3, crossing the factor of cue type (valid vs. neutral), trials were equally divided into three types of cue arrangement. For one third of trials, a single item was cued (*single-cue* trials). For another third of trials, a set of three perceptually grouped items was cued (*grouped-cue* trials). For the rest of trials, a set of random three items was cued (*random-cue* trials). The presence or absence of the multiple-item cueing benefit was examined with cue-set-size-3 trials (i.e., grouped cue, random cue) for the purpose of maximizing the contextual discrepancy

between the cued set of items and the to-be-compared probe in the single-item recognition task (Makovski & Jiang, 2007), but still within the typical VSTM capacity. While the contextual discrepancy between the cued and probed sets of items will be minimal in the set-item recognition task (i.e., three cued items in VSTM, three probed items in the test array), this discrepancy will be maximal in the single-item recognition task (i.e., three cued items in VSTM, a single test probe in the test array). Use of cue set size 3 is consistent with the two studies that originally employed the set-item recognition task (Matsukura et al., 2007; Williams & Woodman, 2012). The design that multiple-item cue trials occupy two thirds of validly cued trials is also consistent with that of Makovski and Jiang (i.e., the ratio of cue-set-size-2 and -3 trials exceeded that of cue-set-size-1 trials).

Third, all experiments were conducted with the articulatory suppression task, in order to encourage the observers to visually maintain the information originally acquired through the visual modality (Luck & Vogel, 2013). While some studies opt to use the color change-detection task to examine the general working memory mechanisms without any articulatory suppression requirement (e.g., p.1079 in Rerko & Oberauer, 2013), we are interested in how visual attention operates within VSTM representations and made efforts toward this goal.

Fourth, because we are interested in how well the observers remember validly cued items relative to neutrally cued items, the observers made an unspeeded manual response (e.g., Makovski et al., 2008; Matsukura et al., 2007, 2014; Matsukura & Hollingworth, 2011; Williams & Woodman, 2012). This method also prevented the observers from prioritizing accuracy over reaction time (RT) for some trials and RT over accuracy for other trials (but see different approaches and methods in Astle et al., 2012; Griffin & Nobre, 2003; Janczyk & Berryhill, 2014; Rerko, Souza, & Oberauer, 2014; Souza, Rerko, & Oberauer, 2014).

Experiment 1

The goal of Experiment 1 was to examine whether attention can select multiple cued items during VSTM maintenance with the set-item recognition task (Fig. 3). The retention-interval cueing benefits were compared when attention was directed to a single item (cue set size 1: single-cue trials) and a set of three items (cue set size 3: grouped-cued trials, random-cue trials) represented in VSTM. The observers were asked to report if there was any color change between the probed set of items in the test array and the cued set of items that had appeared at the corresponding locations in the memory array. If selection of multiple cued items is possible, then a significant cueing benefit should be observed for both cue-set-size-1 and cue-set-size-3 conditions.

Cue-set-size-3 trials were equally divided into *grouped-cue* trials and *random-cue* trials. As mentioned earlier, this manipulation was adopted to examine the effect of perceptual grouping of the cued items on the multiple-item cueing benefit. If the multiple-item cueing benefit can be observed only under the circumstances that attention is oriented with a specific single direction, then a significant cueing benefit should be observed for grouped-cue trials but not for random-cue trials. In contrast, if attention can select multiple cued items positioned on noncontiguous locations during VSTM maintenance, then a significant cueing benefit should be observed for both grouped-cue and random-cue trials.

We first examined the retention-interval cueing benefits with the peripheral dot cues (Exp. 1A) to be comparable with Makovski and Jiang (2007), and then replicated the same pattern of results with the central arrow cues (Exp. 1B). Experiment 1B was conducted to rule out the possibility that the anticipated multiple-item cueing benefit in the random-cue condition of Experiment 1A was driven by relative ease of utilizing peripheral dot cues. Unlike the central arrow cues, peripheral dot cues appear at the exact same perceptual locations with the cued items represented in VSTM. This setting allows the observers to skip interpreting the meaning of the cues before orienting attention to particular locations represented in VSTM (Matsukura et al., 2014; Shimi, Nobre, Astle, & Scerif, 2014).

Experiment 1A

Method

Participants Thirty-two observers participated in Experiment 1A. All observers were University of Iowa undergraduates who participated to receive partial course credits for their involvement; all were between the ages of 18 to 30 years, and all reported having normal or corrected-to-normal visual acuity. None of these observers had participated in any of other experiments reported in the present study.

Stimuli Stimuli were viewed from a distance of 60 cm and were presented on a gray background (22.6 cd/m^2) with a continuously visible white fixation cross (51.5 cd/m^2). The stimuli were presented at six locations that were evenly spaced around an imaginary circle with a radius of 3.8° that was centered at fixation (see Fig. 3). Each memory array consisted of a 1.1° × 1.1° filled square at each of the six locations. The peripheral dot cue was a white, .38° × .38° filled square. The colors were selected equiprobably and randomly from a set of seven easily discriminable colors (without replacement): violet ($x = .245$, $y = .111$, 6.4 cd/m^2), red ($x = .636$, $y = .315$, 12.9 cd/m^2), blue ($x = .152$, $y = .659$, 5.6 cd/m^2), green ($x = .313$, $y = .554$, 20.2 cd/m^2), yellow ($x = .464$, $y = .451$,
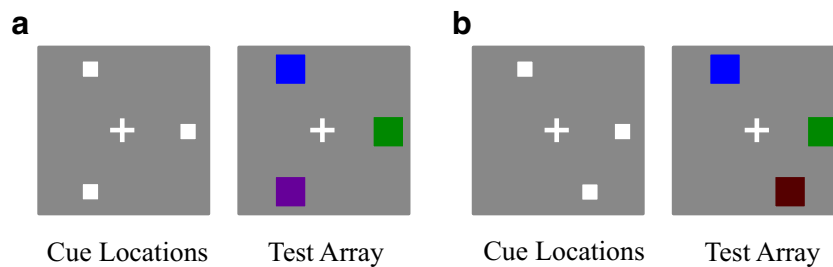
Fig. 4 Two types of random cue arrangement. **a** Each of three dot cues appeared at every other item's location. **b** Two out of three cues appeared at adjacent locations while a single cue appeared in isolation

38.14 cd/m^2), black (x = .299, y = .255, .5 cd/m^2), and brown (x = .582, y = .310, 3.1 cd/m^2). The resulting arrangement with a horizontal axis was identical with the one used in Makovski and Jiang (2007) as well as Matsukura et al. (2014).[3]

Procedure Each trial started with an observer beginning an articulatory suppression task, which the observer was required to repeat either "A, B, C, D" or "1, 2, 3, 4" aloud through the duration of the trial. This concurrent task effectively discourages verbal recoding of visual information (e.g., Baddeley, 1986; Besner, Davies, & Daniels, 1981; Murray, 1968). The observers were instructed to speak at a rate of three or four digits/second or three or four letters/second, and the experimenter continuously monitored the observers to ensure adequate performance.

As illustrated in Fig. 1, the memory array appeared for a duration of 100 ms after a 1,000-ms fixation screen. The offset of the memory array was followed by a blank delay period that was randomly varied within the range from 1,500 to 2,500 ms. Immediately following this delay, the dot cue appeared at one, three, or all six locations that had previously been occupied by color squares. In the single-cue condition, either a single dot cue (validly cued trials, Fig. 3: top row) or six dot cues (neutrally cued trials) appeared for 100 ms. In the grouped-cue condition, either a group of three dot cues on three contiguous locations (validly cued trials, Fig. 3: middle row) or six dot cues (neutrally cued trials) appeared for 100 ms. Three contiguous locations for validly cued trials were randomly selected for each trial. In the random-cue condition, either a set of three dot cues on three randomly selected locations (validly cued

trials, Fig. 3: bottom row) or six dot cues (neutrally cued trials) appeared for 100 ms. Because the current series of experiments were conducted with memory-array set size 6, only two types of random-cue arrangement were possible (Fig. 4). One is that each of three dot cues appeared at every other item's location (Fig. 4a). The other is that two out of three cues appeared at adjacent locations, while a single cue appeared in isolation (Fig. 4b). Both arrangements require the observers to direct their attention to multiple noncontiguous locations (three noncontiguous locations for Fig. 4a, and two noncontiguous locations for Fig. 4b). These two types of random-cue arrangement equiprobably occupied the random-cue trials. The offset of the cue was followed by another blank period that varied randomly within the range from 500 to 1,000 ms. After this delay, a single probe item (the single-cue condition) or a set of three probe items (the grouped-cue condition, the random-cue condition) appeared in the test array and remained on the computer display until an observer made a response. The delay durations between the memory-array offset and the cue onset as well as between the cue offset and the test-array onset were randomly varied to prevent the observers from predicting the presentation timing of the cue and the test array, respectively (e.g., Griffin & Nobre, 2003; Matsukura et al., 2007, 2014). Regardless of whether the retention-interval cue directed the observers' attention to a single item or multiple items, validly cued and neutrally cued trials were equiprobable to increase statistical power and randomized throughout the experiment (e.g., Matsukura et al., 2007, 2014; Matsukura & Hollingworth, 2011).

On half of the trials, the probed set of items was identical to the cued set of items that had been presented on the corresponding locations in the memory array (same-color trials). On the remaining trials, the probed set of items was different from the cued set of items in the memory array by a single item's color (different-color trials, i.e., without replacement). Critically, when a single color change took place in the grouped- and random-cue conditions, (A) this single color change occurred equiprobably to any of three cued locations. (B) Out of three cued locations, the location that a single color change would occur was not predictable for an upcoming trial, as it was randomly determined for each different-color trial. And, (C) the

---

[3] It has been shown that the retention-interval cueing effects remain regardless of whether six items are selected from seven or nine possible colors, using the "without replacement" method (e.g., Makovski, 2012; Makovski & Jiang, 2007; Makovski et al., 2008; Matsukura et al., 2007, 2014). Although it is possible that the observers identify missing colors on each memory array, such identification is equivalent across validly and neutrally cued trials. While this "without replacement" method reduces the likelihood of color-location binding errors or illusory conjunctions (see also p.1078 in Makovski & Jiang, 2007), Williams and Woodman (2012) replicated Matsukura et al.'s set-item recognition results using the "with replacement" method. The equiprobable selection of colors enabled us to balance relative luminance changes for different-color trials (e.g., yellow to black vs. brown to black).

observers were aware of the stipulations of (A) and (B) through the instruction and short practice session. With this design, the observers are required to attend to multiple cued locations, and attending to one out of three cues will mute the cueing benefit (i.e., accuracy of validly cued trials will be driven down to match accuracy of neutrally cued trials).

Using the computer keyboard, the observers pressed "1" if the probed set of items and the cued set of items shared the same colors, and pressed "2" if the probed set of items and the cued set of items differed by a single item's color. The observers made an unspeeded manual response.[4] Each observer participated in a single 60-min experimental session. At the beginning of the session, the observers were given both written and oral instructions. After nine practice trials with the task, each observer completed 420 trials in six blocks of 70 trials. Both cue type (valid vs. neutral) and cue set size (1 [single cue] vs. 3 [grouped cue, random cue]) were randomized throughout each block (thus, throughout the experiment, the mixed-trial design).

Linear decrease of baseline performance Figure 5 illustrates the predicted pattern of results for the set-item recognition task (Fig. 5a) and the single-item recognition task (Fig. 5b), respectively. Because the two tasks predict different patterns of change-detection performance (not only for validly cued trials but also) for neutrally cued trials, the logic behind these differences should be clearly laid out before presenting the results of the set-item recognition experiments (Exp. 1).

During sensory processing, perceptual qualities of the attended items decline as set size of the attended items increases, because a finite neural resource has to spread out (e.g., Anderson et al., 2013; Bundesen, 1990; Palmer, Fencsik, Flusberg, Horowitz, & Wolfe, 2011; Posner, Snyder, & Davidson, 1980). If the sensory recruitment hypothesis is true, then accuracy of validly cued trials should decline as cue set size increases (e.g., Mazyar, van den Berg, & Ma, 2012; Palmer, 1990, for the set size effects on visual search during VSTM maintenance).[5]

In the present study, given perceptual grouping of multiple cued items is manipulated *within* the cue-set-size-3 condition, we cannot expect a dramatic linear decrease of

---

[4] All RT analyses of correct trials in the present study yielded the same pattern (shorter RT for validly cued trials than for neutrally cued trials) as the analyses of percent correct. When RTs shorter than 150 ms or longer than 2,000 ms were excluded from the analyses, less than 6.0 % of correct trials were removed. Median RT analyses without trimming produced the same pattern of results.

[5] We acknowledge that the current set-item recognition experiment does not have any means to assess representational qualities of the cued and un-cued colors, because it is a recognition task that involves location binding. Accordingly, the results from Experiment 1 should not be confused with the results from the color recall studies (e.g., Hollingworth & Hwang, 2013; Souza, Rerko, Lin, & Oberauer, 2014; Williams et al., 2013). These studies estimated precision of the cued and uncued colors, dissociated from the retention probabilities of the cued and uncued items.

a Predicted Baseline: Set Item Recognition Experiment 1A
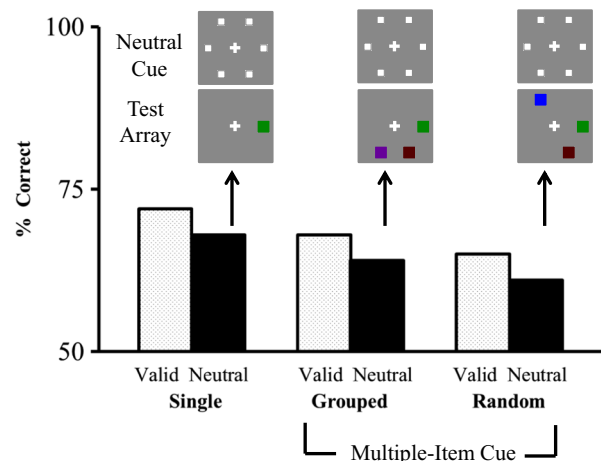
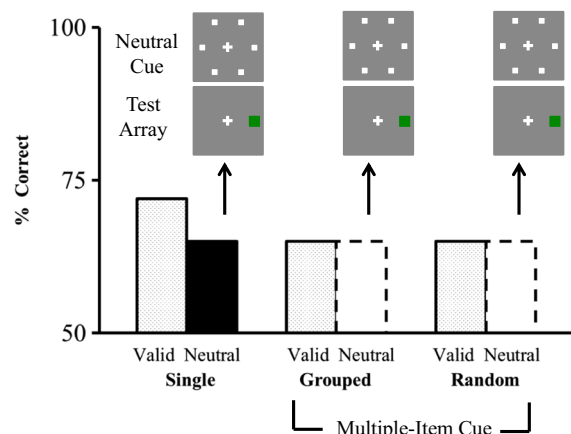**b Predicted Baseline: Single Item Recognition Based on Makovski and Jiang (2007)**

Fig. 5 **a** Predicted change-detection accuracy pattern of neutrally cued trials with the set-item recognition task employed in Experiment 1A. The *x*-axis represents cue type (valid, neutral) and cue set size (1 [single] vs. 3 [grouped, random]). **b** Predicted change-detection accuracy pattern of neutrally cued trials with the single-item recognition task based on Makovski and Jiang (2007)

accuracy as perceptual grouping of the cued items is removed. However, if memory-level attention operates in similar manners with perception-level attention within the typical VSTM capacity, then change-detection accuracy of validly cued trials should be higher for grouped-cue trials than for random-cue trials. Spatially-directed attention is known to be sensitive to perceptual grouping of multiple attended items during sensory processing (e.g., Egly, Driver, & Rafal, 1994; Hecht & Vecera, 2007; Matsukura & Vecera, 2006; see Woodman, Vecera, & Luck, 2003, for the analogous effect on VSTM encoding).

Likewise, accuracy of neutrally cued trials should decline as cue set size increases and perceptual grouping of the cued

items is removed (see Fig. 5a). This is because, in the set-item recognition task, the test arrays of neutrally cued trials are held constant with the test arrays of validly cued trials; regardless of whether a trial is validly or neutrally cued, at the time of the recognition test, the observers will face a single probe item in the cue-set-size-1 condition and a set of three probe items in the cue-set-size-3 condition. In contrast, in the single-item recognition task (Fig. 5b, the prediction based on Makovski & Jiang, 2007), neutrally cued trials share a single test probe across different cue-set-size conditions. Makovski and Jiang compared change-detection accuracy of validly cued trials in each cue-set-size condition against this "shared" neutrally cued condition (in their study, no-cue trials). Here, the question of "whether or not set-item recognition accuracy of neutrally cued trials (Fig. 5a) should be collapsed across different cue set sizes" arises (as seen in Makovski and Jiang's single-item recognition task). In case of the set-item recognition task, collapsing accuracy of neutrally cued trials across different cue-set-size conditions will unnecessarily enlarge the retention-interval cueing benefit in the single-cue condition by lowering mean accuracy of neutrally cued trials. At the same time, such averaging will also shrink the size of the retention-interval cueing benefit in the random-cue condition by raising mean accuracy of neutrally cued trials. To avoid such artificial masking effects, we presented accuracy of neutrally cued trials for each cue set size (and cue arrangement) separately (Fig. 6). Indeed, anticipated mean accuracy of neutrally cued trials linearly declined across single-, grouped-, and random-cue conditions of Experiment 1, $F(2, 94) = 25.23$, $p < .0001$, $\eta_p^2 = .35$.[6] Even within the cue-set-size-3 condition, a significant linear decrease was observed as perceptual grouping of neutrally cued items was removed, $F(1, 47) = 13.29$, $p < .001$, $\eta_p^2 = .22$.

Results and discussion

Figure 6a shows mean change-detection accuracy (percent correct, collapsed across same-color and different-color trials) from Experiment 1A as a function of cue type (valid vs. neutral) and cue set size (1 [single cue] vs. 3 [grouped cue, random cue]). To rule out possible distortions from response bias, all the data in the present study were also analyzed with d', a measure of sensitivity based on the signal detection theory (Macmillan & Creelman, 1991; see Appendix).[7] Because the analyses of d' yielded the same pattern of results

---

[6] None of the repeated-measures ANOVA in the present study violated the sphericity assumption, $ps > .1$.

[7] Because the retention-interval cueing paradigm essentially violates the critical assumption that the observers try to maintain as many items as possible in VSTM (e.g., Rouder, Morey, Morey, & Cowan, 2011) by requiring the observers to remember (i.e., preferentially retain) either a single cued item (cue set size 1) or the cued set of items (cue set size 3), it is not appropriate to estimate the capacity for validly cued trials (Matsukura et al., 2014; Matsukura & Hollingworth, 2011).
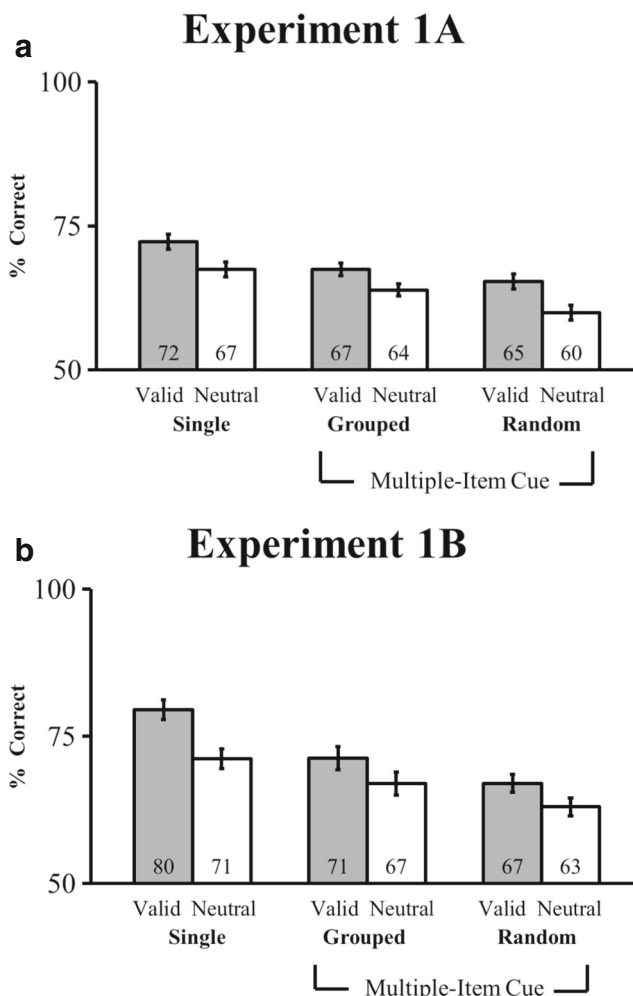

Fig. 6 a Mean change-detection accuracy from Experiment 1A (set-item recognition) as a function of cue type (valid, neutral) and cue set size (1 [single] vs. 3 [grouped, random]). b Mean change-detection accuracy from Experiment 1B as a function of cue type (valid, neutral) and cue set size (1 [single] vs. 3 [grouped, random]). For this and all subsequent figures, error bars represent 95 % within-subjects confidence intervals (Loftus & Masson, 1994)

as the analyses of percent correct, we will focus on accuracy for the rest of the study. For all cue-set-size conditions, the observers recognized validly cued items more accurately than neutrally cued items (i.e., the retention-interval cueing benefit). Given the goal of Experiment 1A was to examine whether attention can select multiple cued items during VSTM maintenance or not, we will compare the cueing benefits across the single-item cueing condition (cue set size 1: single-cue trials) and the multiple-item cueing condition (cue set size 3: grouped-cue trials, random-cue trials) first, and then move onto examining the effect of perceptual grouping on the multiple-item cueing benefit observed within the cue-set-size-3 condition (grouped-cue trials vs. random-cue trials).

An analysis of variance (ANOVA) with within-subjects factors of cue type (valid vs. neutral) and cue set size (1 vs. 3)

was conducted. Higher accuracy in validly cued trials than in neutrally cued trials led to a significant main effect of cue type, $F(1, 31) = 45.03$, $p < .0001$, $\eta_p^2 = .59$. Higher accuracy in cue-set-size-1 trials than in cue-set-size-3 trials also led to a significant main effect of cue set size, $F(1, 31) = 33.76$, $p < .0001$, $\eta_p^2 = .52$. Despite overall accuracy was higher in cue-set-size-1 trials than in cue-set-size-3 trials, the two-way interaction of cue type and cue set size was not significant, $F(1, 31) = .04$, $p = .85$, $\eta_p^2 = .00$, as the size of the cueing benefit was approximately the same between cue-set-size-1 and -3 trials.

Having observed significant cueing benefits in both cue-set-size-1 and -3 conditions, we now move onto examining the effect of perceptual grouping on the multiple-item cueing benefit by an ANOVA with within-subjects factors of cue type and perceptual grouping (grouped-cue trials vs. random-cue trials). This planned within-cue-set-size-3-condition analysis allowed us to determine whether selection of multiple cued items is possible only under the circumstances that attention is oriented to a single perceptually-grouped set of items with a specific direction or not. Higher accuracy in validly cued trials than in neutral cued trials led to a significant main effect of cue type, $F(1, 31) = 29.73$, $p < .0001$, $\eta_p^2 = .49$. Slightly higher accuracy in grouped-cue trials than in random-cue trials led to a significant main effect of perceptual grouping, $F(1, 31) = 9.94$, $p < .004$, $\eta_p^2 = .24$. The two-way interaction of cue type and perceptual grouping was not significant, $F(1, 31) = 1.18$, $p = .29$, $\eta_p^2 = .04$, as the magnitude of the cueing benefit was approximately equivalent between grouped- and random-cue trials.

Planned pair-wise comparisons confirmed that the observed cueing benefit was significant for single-cue trials, $t(31) = 3.78$, $p < .0007$, for grouped-cue trials, $t(31) = 3.44$, $p < .002$, and for random-cue trials, $t(31) = 4.21$, $p < .0002$, respectively. The multiple-item cueing benefit observed for grouped-cue trials successfully replicated the results originally reported by Matsukura et al. (2007) as well as Williams and Woodman (2012). Having observed a significant cueing benefit for random-cue trials, some may wonder whether attention was directed differently when three dot cues appeared on three noncontiguous locations (Fig. 4a) and when two out of three cues appeared on contiguous locations (Fig. 4b). Planned pair-wise comparison confirmed no difference in change-detection accuracy between these two types of random-cue arrangement (identical 65.2 %), $t(31) = .00$, $p = 1.00$. These results rule out the possibility that the cueing benefit observed in the random-cue condition was driven by a particular type of random-cue arrangement. Together, Experiment 1A results suggest that, regardless of whether multiple cued items are grouped into a single perceptual group or not, attention can select multiple cued items during VSTM maintenance.

Experiment 1B

Because Matsukura et al. (2014) recently demonstrated that the peripheral dot cue used in Experiment 1A can guide attention during VSTM maintenance as efficiently as the central arrow cue, the possibility that the cueing benefits observed in Experiment 1A are generated through methodological artifacts such as interruption masking or corrective eye movements (that are more likely to occur in neutrally cued trials) is ruled out. However, one concern for Experiment 1A results is that a significant multiple-item cueing benefit observed for random-cue trials might have been caused by relative ease of utilizing peripheral dot cues. Because peripheral dot cues appear at exactly the same perceptual locations with the cued items represented in VSTM, the observers are not required to interpret the meaning of the three cues before they direct attention to multiple noncontiguous locations. Consequently, it is easier for the observers to process a shape formed by three dot cues. To rule out such a possibility, in Experiment 1B, we replicated Experiment 1A using the central arrow cues, which prevented the observers from easily forming a shape perception of three cues. If the multiple-item cueing benefit observed in the random-cue condition of Experiment 1A pertains to use of peripheral dot cues per se, then no cueing benefit should be observed in the random-cue condition of Experiment 1B. In contrast, if the multiple-item cueing benefit in the random-cue condition of Experiment 1A reflected a sound operation of memory-level attention broadly, then the cueing benefit comparable to Experiment 1A should be observed in the random-cue condition of Experiment 1B.

Method

The method of Experiment 1B was identical to that of Experiment 1A except (1) a new group of 16 observers participated in the experiment, and (2) the dot cue was replaced with the central cue, which was a white arrow (51.5 cd/m²), 1.9° in length (Matsukura et al., 2007, 2014). In validly cued trials for the single-cue condition, a single central arrow pointed to one of the six locations that were previously occupied by six color squares. In validly cued trials for the grouped-cue condition, a group of three arrows pointed to three contiguous locations. In validly cued trials for the random-cue condition, three arrows pointed to three randomly selected locations. In neutrally cued trials, the cue was a set of six arrows pointing to each of the six locations regardless of cue set size/cue arrangement. Because the main results of Experiment 1A were significant with 16 or fewer observers, data collection ceased at 16 observers.

Results and discussion

Figure 6b shows mean change-detection accuracy from Experiment 1B as a function of cue type and cue set size. While overall accuracy experienced mild increase relative to Experiment 1A (regardless of whether equal variances are assumed or not, $ps > .10$), the cueing benefits were replicated across the single-, grouped-, and random-cue conditions, and these observations were supported by an ANOVA with within-subjects factors of cue type and cue set size. Although it is not as detrimental as driving accuracy of validly cued trials down to match that of neutrally cued trials, the tendency that use of the peripheral cue slightly lowers mean change-detection accuracy relative to use of the central cue is well documented (e.g., Pertzov et al., 2013; Shimi et al., 2014; also reviewed in Matsukura et al., 2014). Higher accuracy in validly cued trials than in neutrally cued trials led to a significant main effect of cue type, $F(1, 15) = 38.97$, $p < .0001$, $\eta_p^2 = .72$. Higher accuracy in cue-set-size-1 trials than in cue-set-size-3 trials also led to a significant main effect of cue set size, $F(1, 15) = 35.84$, $p < .0001$, $\eta_p^2 = .71$. A slightly larger cueing benefit in cue-set-size-1 trials than in cue-set-size-3 trials produced a significant two-way interaction of cue type and cue set size, $F(1, 15) = 5.45$, $p < .03$, $\eta_p^2 = .27$.

Having replicated significant cueing benefits in both cue-set-size-1 and -3 conditions, we now move onto examining the effect of perceptual grouping on the multiple-item cueing benefit by an ANOVA with within-subjects factors of cue type and perceptual grouping. Higher accuracy in validly cued trials than in neutrally cued trials led to a significant main effect of cue type, $F(1, 15) = 14.55$, $p < .002$, $\eta_p^2 = .49$. Higher accuracy in grouped-cue trials than in random-cue trials led to a significant main effect of perceptual grouping, $F(1, 15) = 11.55$, $p < .004$, $\eta_p^2 = .44$. Replicating Experiment 1A results, the two-way interaction of cue type and perceptual grouping was not significant, $F(1, 15) = .02$, $p = .88$, $\eta_p^2 = .00$, as the magnitude of the cueing benefit was approximately equivalent between grouped- and random-cue trials.

Planned pair-wise comparisons confirmed that the observed cueing benefit was significant for single-cue trials, $t(15) = 5.38$, $p < .0001$, for grouped-cue trials, $t(15) = 2.39$, $p < .03$, and for random-cue trials, $t(15) = 2.83$, $p < .01$, respectively. Replicating Experiment 1A results, no change-detection accuracy difference was found between the valid random-cue trials that each of three arrow cues pointed to every other item's location (68 %, similar to Fig. 4a) and those that two out of three arrow cues pointed to adjacent locations (65 %, similar to Fig. 4b), $t(15) = 1.29$, $p = .22$. Together, Experiment 1B results suggest that the multiple-item cueing benefit observed in the random-cue condition of Experiment 1A was not driven by relative ease of utilizing peripheral dot cues. Consistent with Matsukura et al. (2007, 2014), these results also ruled out the possibility that the cueing benefits

observed in Experiment 1A were caused by interruption masking or corrective eye movements that are more likely to occur in neutrally cued trials.

Experiment 2

Having demonstrated that attention can indeed select multiple cued items during VSTM maintenance when the observers are asked to recognize a set of items in the manner that these items were originally attended, we now move on to replicate the experiment with the single-item recognition task (Fig. 7). As mentioned in Introduction, in order for us to argue that multiple retention-interval cues can be efficiently utilized when there is a reasonable contextual match between the cued representations in VSTM and the to-be-compared percepts in the test array (set-item recognition), it is necessary to show the other side of this argument. That is, lack of the multiple-item cueing benefit (originally reported by Makovski & Jiang, 2007) has to be successfully replicated when there is a discernible contextual discrepancy between the cued representations in VSTM
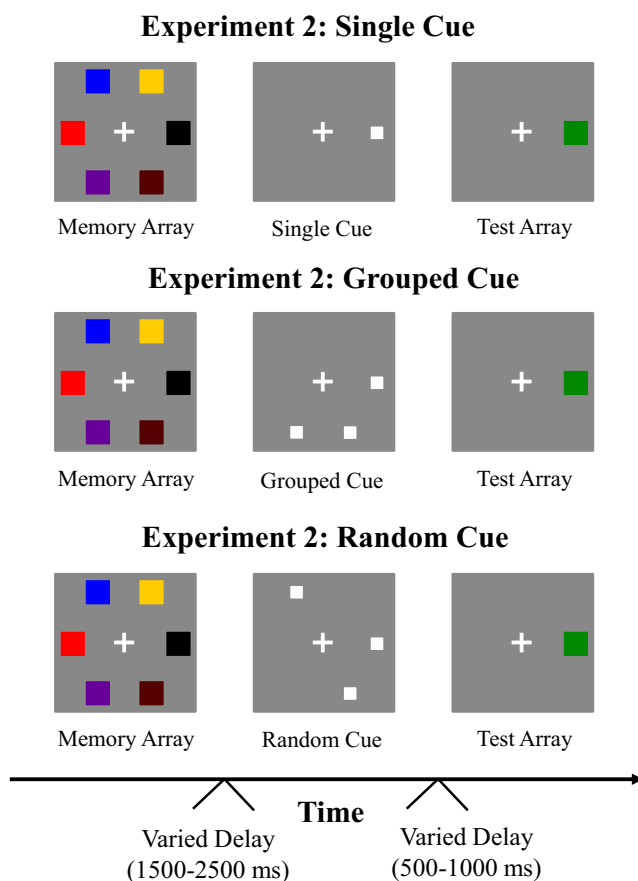
Experiment 2: Single Cue

Memory Array   Single Cue   Test Array

Experiment 2: Grouped Cue

Memory Array   Grouped Cue   Test Array

Experiment 2: Random Cue

Memory Array   Random Cue   Test Array

Time

Varied Delay (1500-2500 ms)   Varied Delay (500-1000 ms)

Fig. 7 Event sequence of a valid retention-interval cue trial for the single (top), grouped (middle), and random (bottom) cue conditions in Experiment 2 (the single-item recognition task)

and the to-be-compared percept in the test array (single-item recognition).

If the observers are able to use multiple retention-interval cues when no discernible contextual discrepancy exists between the cued set of items and the probed set of items, then a significant cueing benefit should be observed in the cue-set-size-1 condition, but not in the cue-set-size-3 condition. In the cue-set-size-1 condition, there will be no contextual discrepancy between the cued and probed items (i.e., a single cued item in VSTM, a single test probe in the test array). However, in the cue-set-size-3 condition, there will be a discernible contextual discrepancy between the cued set of items in VSTM and the probed item (i.e., three cued items in VSTM, a single test probe in the test array).

While it is not a typical procedure to conduct the retention-interval cueing experiment, in order to examine whether memory-level attention is "qualitatively" different from perception-level attention, Makovski and Jiang (2007, their Exp. 2A) conducted their single-item recognition experiment by mixing in precue trials for half of entire trials (i.e., the mixed-trial design experiment with precue trials). That is, while the observers were performing the retention-interval cueing task for 50 % of the time, they were also performing the precueing task for the other 50 % of the time. As Schmidt, Vogel, Woodman and Luck (2002) demonstrated that, when an attention-directing cue is presented before the memory array onset or immediately after the memory array offset, attention plays a role of transferring perceptual representations of the cued item into VSTM (Averbach & Coriell, 1961; Griffin & Nobre, 2003; Woodman et al., 2003; see Awh & Pashler, 2000; Ester et al., 2014; Hahn & Kramer, 1998; Kramer & Hahn,

1995, for multiple-item precueing benefits; see Sperling, 1960, for the perceptually grouped multiple-item cueing benefit within the iconic memory range). Here, to be comparable with Makovski and Jiang, we first replicated the single-item recognition experiment with the mixed-trial design (i.e., 50 % precued trials). Fully acknowledging that the design of Experiment 2 is not comparable with that of Experiment 1, we will examine whether the same pattern of results can be replicated with the standard retention-interval cueing procedure (without precue trials) in Experiments 3 and 4.

Method

The method was identical to that of Experiment 1A except (1) a new group of 32 observers participated in the experiment. (2) At the time of the test, the observers were asked to report whether a single test item had the same or different color from the cued item appeared at the corresponding location in the memory array. For cue-set-size-3 trials, a single test probe always appeared at one of three cued items' locations (single-item recognition, Fig. 7). As in the set-item recognition task (Exp. 1), when a single color change took place in the cue-set-size-3 condition, this change occurred equiprobably to any of three cued locations. (3) Crossing cue type (valid vs. neutral) and cue set size (1 [single cue] vs. 3 [grouped cue, random cue]), half of trials were precued (Fig. 8). Because the experimental design that includes precue trials for 50 % of the time reduces the number of retention-interval cue trials into half of the previous experiments, the number of observers was brought back to 32 observers; however, the main results were significant
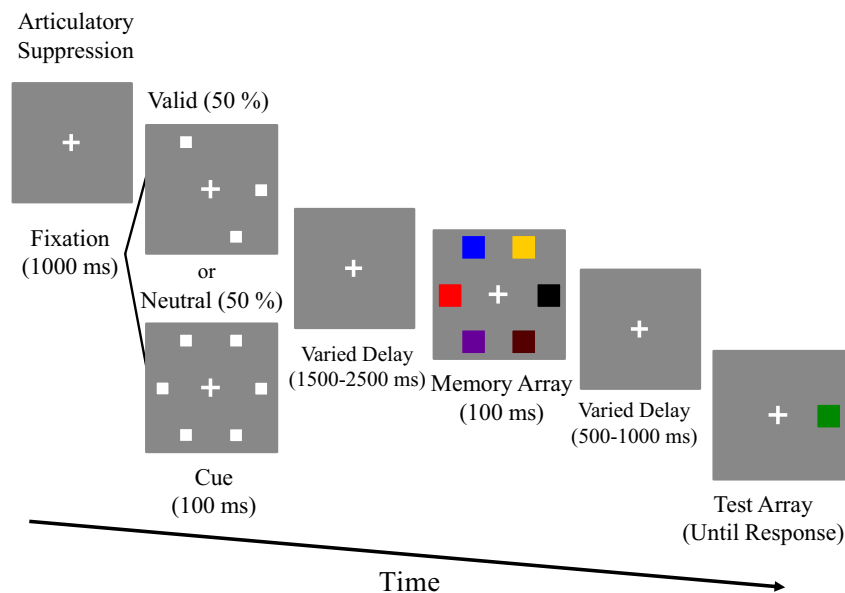

Fig. 8 Event sequence of a precue trial (the random-cue condition) in Experiment 2 (the single-item recognition task). The durations of the whole trial as well as the retention interval were equated with those of a retention-interval cue trial (Fig. 1) that occupied half of the entire experiment

with 16 or fewer observers. An increase in the number of observers was also to be comparable with the corresponding experiment in Makovski and Jiang (2007, Exp. 2A, $n = 23$). (4) The durations of the whole trial, each delay, and exposure duration of stimuli were equivalent between precue and retention-interval cue trials. This equivalence was achieved by simply switching the presentation timing of the memory array and the cue (compare Figs. 1 and 8). (5) Consistent with Makovski and Jiang, precue and retention-interval cue trials were randomized throughout each block (thus, throughout the experiment, the mixed-trial design).

No linear decrease of baseline performance To be comparable with Experiment 1, a single test probe of each trial was created by actually removing the other two items from the probed set of three items of each trial used in Experiment 1A. This tracking method allowed us to examine whether accuracy of neutrally cued trials in Experiment 2 would

a Pre-Cue Trials

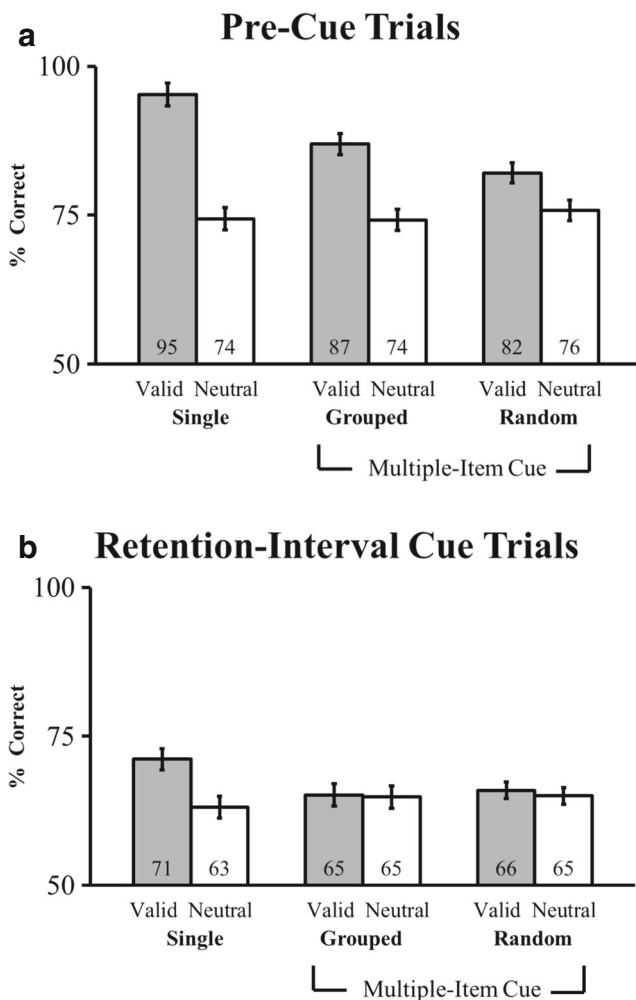**b** Retention-Interval Cue Trials

Fig. 9 Experiment 2 (single-item recognition) results: **a** Mean change-detection accuracy of precue trials as a function of cue type (valid, neutral) and cue set size (1 [single] vs. 3 [grouped, random]). **b** Mean change-detection accuracy of retention-interval cue trials as a function of cue type (valid, neutral) and cue set size (1 [single] vs. 3 [grouped, random])

be approximately equal across the single-, grouped-, and random-cue conditions (predicted in Fig. 5b), as opposed to a linear decrease observed in Experiment 1 (Fig. 6). Indeed, no significant linear decrement of accuracy was found for neutrally cued trials across the single-, grouped-, and random-cue conditions of Experiment 2, $F(2, 62) = .07$, $p = .92$, $\eta_p^2 = .00$, in the precue condition, and $F(2, 62) = .66$, $p = .52$, $\eta_p^2 = .02$, in the retention-interval cue condition, respectively. This pattern persisted for the rest of the single-item recognition experiments in the present study ($ps > .5$).

Results and discussion

Figure 9 illustrates mean change-detection accuracy from Experiment 2 as a function of cue type and cue set size. Replicating the previous reports (e.g., Ester et al., 2014), the pre-cueing benefits were observed regardless of cue set size (Fig. 9a). And, of our interest, replicating Makovski and Jiang (2007), a significant retention-interval cueing benefit was observed for the cue-set-size-1 condition, but not for the cue-set-size-3 condition (Fig. 9b). To be comparable with Makovski and Jiang's analysis format, we will analyze precue trials first, and then move onto the analysis of retention-interval cue trials by an ANOVA with within-subjects factors of cue type and cue set size.

Precueing benefit Higher accuracy in validly cued trials than in neutrally cued trials led to a significant main effect of cue type, $F(1, 31) = 124.28$, $p < .0001$, $\eta_p^2 = .80$. Higher accuracy in cue-set-size-1 trials than in cue-set-size-3 trials also led to a significant main effect of cue set size, $F(1, 31) = 58.59$, $p < .0001$, $\eta_p^2 = .65$. A larger cueing benefit in cue-set-size-1 trials than in cue-set-size-3 trials produced a significant two-way interaction of cue type and cue set size, $F(1, 31) = 40.02$, $p < .0001$, $\eta_p^2 = .56$.

Having replicated significant precueing benefits in both cue-set-size-1 and cue-set-size-3 conditions, we now move onto examining the effect of perceptual grouping on the multiple-item cueing benefit by an ANOVA with within-subjects factors of cue type and perceptual grouping. Higher accuracy in validly cued trials than in neutrally cued trials led to a significant main effect of cue type, $F(1, 31) = 56.96$, $p < .0001$, $\eta_p^2 = .65$. No main effect of perceptual grouping was observed, $F(1, 31) = 2.39$, $p = .13$, $\eta_p^2 = .07$. However, the two-way interaction of cue type and perceptual grouping was significant, $F(1, 31) = 6.09$, $p < .02$, $\eta_p^2 = .16$, because the size of the precueing benefit was slightly larger in grouped-cue trials than in random-cue trials.

Planned pair-wise comparisons confirmed that the observed precueing benefit was significant for single-cue trials, $t(31) = 11.07$, $p < .0001$, for grouped-cue trials, $t(31) = 7.34$, $p < .0001$, and for random-cue trials, $t(31) = 4.21$, $p < .0002$, respectively. Replicating the previous reports (e.g., Awh &

Pashler, 2000), these results suggest that attention can select multiple cued items and encode perceptual representations of these selected items into VSTM, when the attention-directing cues appear before the presentation of to-be-remembered items. Moreover, replicating Makovski and Jiang (2007; see also Ester et al., 2014), when attention influences the encoding process (as opposed to the maintenance process), the visual system seems to be able to overcome the contextual discrepancy emerged between the cued set of items in VSTM and a single test probe in perception. Significant multiple-item precueing benefits were observed even with the single-item recognition task.

Retention-interval cueing benefit Higher accuracy in validly cued trials than in neutrally cued trials led to a significant main effect of cue type, $F(1, 31) = 17.09, p < .0001, \eta_p^2 = .35$. No main effect of cue set size was observed, $F(1, 31) = 2.50, p = .12, \eta_p^2 = .08$. However, the presence of the cueing benefit in cue-set-size-1 trials (but not in cue-set-size-3 trials) led to a significant two-way interaction of cue type and cue set size, $F(1, 31) = 14.36, p < .001, \eta_p^2 = .32$. As evident in Fig. 9b, the planned within-cue-set-size-3 analyses showed neither the main effects of cue type, $F(1, 31) = .40, p = .53, \eta_p^2 = .01$, perceptual grouping, $F(1, 31) = .11, p = .74, \eta_p^2 = .00$, nor the interaction of cue type and perceptual grouping, $F(1, 31) = .05, p = .83, \eta_p^2 = .00$.

Planned pair-wise comparisons confirmed that the observed retention-interval cueing benefit was significant for single-cue trials, $t(31) = 4.55, p < .0001$, but neither for grouped-cue trials, $t(31) = .19, p = .85$, nor for random-cue trials, $t(31) = .67, p = .51$. Replicating Makovski and Jiang's (2007) results, the single-item recognition experiment produced a significant single-item cueing benefit but not the multiple-item cueing benefit during VSTM maintenance. These results suggest that, when attention influences the maintenance process (as opposed to the encoding process), the visual system cannot survive the contextual discrepancy arisen between the cued set of items in VSTM and a single test probe in perception. Multiple-item retention-interval cueing benefits were abolished with the single-item recognition task.

The interaction of cue presentation timing and cue set size In order to examine whether selection of multiple cued items is "qualitatively" different between precue and retention-interval cue trials, Makovski and Jiang (2007, p. 1076) conducted an ANOVA with the cue presentation timing (precue vs. retention-interval cue) and cue set size (2 vs. 6), and observed a significant interaction. While the exact reason for why the two levels of cue set size were set to 2 and 6 was not reported, this analysis corresponds to a significant two-way interaction of cue presentation timing (precue vs. retention-interval cue) and cue set size (3 [grouped, random] vs. 6 [neutral cues]) observed in the present study, $F(1, 31) = 46.10, p < .0001, \eta_p^2 = .60$. Of importance, the presence of multiple-item *precueing* benefits also rules out the possibility

that lack of the multiple-item *retention-interval* cueing benefit is attributed to decision noise. If lack of the multiple-item retention-interval cueing benefit is caused by the increased number of cue locations per se, then no multiple-item precueing benefits should have been observed.

The goal of the present study was to investigate whether attention can select multiple cued items during VSTM maintenance when the observers are asked to recognize a set of items in the manner that these items were originally attended, and the results from Experiments 1 and 2 complete both sides of our theoretical argument. On the one side, attention can select more than a single item as long as there is a reasonable contextual match between the cued representations in VSTM and the to-be-compared percepts (Exp. 1). On the other side, attention cannot select multiple cued items during VSTM maintenance when a single test probe in the test array carries a clear contextual discrepancy with the cued set of items represented in VSTM (Exp. 2).

While the mixed-trial design experiment with 50 % precue trials accomplished the goal of Makovski and Jiang's (2007) study, such a design has not been a typical procedure used to examine the nature of memory-level attention. Indeed, single-item retention-interval cueing effects, which were originally reported with the mixed-trial design that included precue trials (Griffin & Nobre, 2003), have been successfully replicated with retention-interval cue trials *alone* a number of times (as reviewed in Introduction). Given such a record in the literature, it was critical for us to confirm that the pattern of results observed in Experiment 2 is generalizable to when the experiment is conducted without any precue trials. Experiment 3 was conducted to achieve this goal.

Experiment 3

The goal of Experiment 3 was to simply replicate the single-item recognition task results observed in Experiment 2 without any precue trials. Critically, its design was identical to that of Experiment 1A except the observers were asked to recognize a single item out of three cued items. If Experiment 2 results are generalizable to the standard retention-interval cueing paradigm, then a significant cueing benefit should be observed in the cue-set-size-1 condition but not in the cue-set-size-3 condition.

Method

The method was the same as that of Experiment 2 except (1) a new group of 16 observers participated in the experiment, and (2) precue trials were removed and replaced with retention-interval cue trials. To reiterate, the resulting design was identical to that of Experiment 1A except the task was switched from set-item recognition to single-item recognition.

Results and discussion

Figure 10 shows mean change-detection accuracy from Experiment 3 as a function of cue type and cue set size. Astonishingly, all retention-interval cueing benefits were abolished. Indeed, an ANOVA with within-subjects factors of cue type and cue set size indicated that neither the main effect of cue type, $F(1, 15) = .00$, $p = .99$, $\eta_p^2 = .00$, nor the interaction of cue type and cue set size was significant, $F(1, 15) = 2.4$, $p = .14$, $\eta_p^2 = .14$. Slightly higher accuracy in cue-set-size-1 trials than in cue-set-size-3 trials produced a trend toward significance for the main effect of cue set size, $F(1, 15) = 3.54$, $p < .08$, $\eta_p^2 = .19$.

Admittedly, the results of Experiment 3 were a surprise; not even a single-item cueing benefit was observed. If attention can select multiple cued items during VSTM maintenance as long as there is a reasonable contextual match between the cued representations in VSTM and the to-be-compared percepts in the test array, then a significant cueing benefit in the cue-set-size-1 condition should have been replicated without any precue trials.

Why did the single-item cueing benefit go away when the single-item recognition experiment was replicated with retention-interval cue trials alone? One strong possibility for the complete disappearance of the single-item cueing benefit in Experiment 3 is that the observers stopped using the retention-interval cueing information, as soon as they figured that such information was not useful to perform a given color change-detection task for the majority of trials. As briefly touched in Introduction, during sensory processing, the exogenous cueing benefit is generated through the mechanism that the observer's attention is unpredictably drawn to the peripherally cued location by virtue of luminance transients (e.g., Jonides, 1981; Posner & Cohen, 1984). Indeed, the automatic

Experiment 3

Fig. 10 Mean change-detection accuracy from Experiment 3 (single-item recognition) as a function of cue type (valid, neutral) and cue set size (1 [single] vs. 3 [grouped, random])

nature of the exogenous cue is usually examined by lowering the cue validity below 50 %. However, the exogenous (peripheral) retention-interval cue used in the present study was presented long after an iconic image of to-be-remembered items had faded away, and a single test probe appeared, again, long after an icon of the cue had disappeared (beyond the iconic memory range, Irwin & Yeomans, 1986; Sperling, 1960). Interestingly, a recent work conducted by Shimi et al. (2014) demonstrated that, when the retention-interval cue validity was reduced to 25 % of entire trials, the cueing benefit and cost were completely eliminated, regardless of whether the cue was presented peripherally or centrally. Based on these results, Shimi et al. concluded that, unlike the exogenous cueing benefit observed during sensory processing, attention guided by the exogenous cue during VSTM maintenance operates through the goal-directed selection (thus, not qualified as the bona fide exogenous cue).

If memory-level attention operates in a manner that is sensitive to the observers' goal and motivation, then it is sensible to hypothesize that all retention-interval cueing benefits observed in the present study were generated through the goal-directed selection. Then, what exactly separated the results of Experiment 2 from those of Experiment 3? The single-item cueing benefit was observed only when precue trials were mixed in the experiment (Exp. 2), but not when the experiment was composed of retention-interval cue trials alone (Exp. 3). For both experiments, the cue validity (valid vs. neutral) was held constant at 50 %. However, in Experiment 2, besides 50 % of entire validly cued trials were precued, 16.7 % of entire validly cued trials were cued by a single retention-interval cue, which the probed and cued items contextually matched. That is, reminiscent to the standard cue validity increase, the majority of attention-directing cues (66.7 %) were "useful" to perform the color change-detection task.[8] Consistent with the previous reports (e.g., Ester et al., 2014; Makovski & Jiang, 2007), when attention influences the encoding process (precue trials), the visual system appears to be able to overcome the contextual discrepancy between the cued set of items in VSTM and a single test probe in perception. And, for single-cue trials, regardless of whether a trial was precued or cued during the retention interval, the observers were aware that there was no contextual discrepancy between the cued and probed items. Based on such a structure of the experiment, we hypothesize that the observers in Experiment 2 were motivated to utilize the cueing information.

In contrast, in Experiment 3, the majority of valid cues were "not so useful" to perform the task. Two thirds of validly cued trials (66.7 %) suffered from the contextual
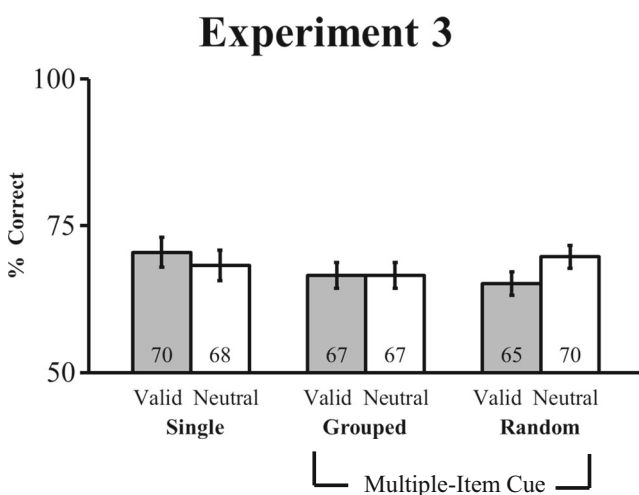
---

[8] In case of Makovski and Jiang's (2007) multiple dot cue experiment (Exp. 2A), besides 50 % of entire validly cued trials were precued, 10 % of entire validly cued trials were cued by a single retention-interval cue, which the probed and cued items contextually matched.

mismatch between the cued set of items in VSTM and a single test probe (cue-set-size-3 trials). The proportion of validly cued trials that did not suffer from the contextual mismatch between the cued and probed items was reduced to only 33.3 % of entire validly cued trials (cue-set-size-1 trials). If memory-level selection is sensitive to the observers' goal and motivation, then it is not difficult to imagine that the observers were discouraged to utilize any retention-interval cueing information and simply start ignoring the cue. Both cue type (valid vs. neutral) and cue set size (1 [single cue] vs. 3 [grouped cue, random cue]) were equiprobably randomized throughout each block (see the Method section of Exp. 1A).

Given such an apparent structural difference between Experiments 2 and 3, it was imperative for us to examine whether the disappearance of the single-item cueing benefit in Experiment 3 is attributed to the goal-directed nature of memory-level selection, but with a method other than increasing the number of trials with the useful cue (as seen in Exp. 2). While it is easy for us to merely increase the proportion of single-cue trials within Experiment 3 to test the posited hypothesis, an analogous design was already employed in Experiment 2 (see also Shimi et al., 2014, for the cue validity increase), and this method reduces statistical power of multiple-item cueing trials. If possible, the hypothesis that the disappearance of the single-item cueing benefit in Experiment 3 derives from the goal-directed nature of memory-level selection should be tested while holding the ratio between cue-set-size-1 and -3 trials constant with the previous experiments. Experiment 4 was conducted to achieve this goal by *dissociating* the observers' motivation to use the cueing information for cue-set-size-1 and -3 trials, while keeping the ratio of cue-set-size-1 and -3 trials identical to Experiment 3.

Experiment 4

The goal of Experiment 4 was simply to bring back the single-item cueing benefit by motivating the observers to use the cueing information for cue-set-size-1 trials (but not for cue-set-size-3 trials). In Experiment 4, we simply replicated Experiment 3 but blocked trials by the single-, grouped-, and random-cue conditions. The observers were explicitly told which cue arrangement type (single, grouped, random) they would encounter at the beginning of every two blocks. As described in the Method section of Experiment 1A, all experiments in the present study were conducted with six blocks for short breaks. This blocked-trial design enabled us to re-create the environment that the observers are encouraged to use the single retention-interval cue without increasing the number of trials with the useful cue (e.g., throwing in precue trials).

Of importance, the cue validity (valid vs. neutral) remained at 50 % as in all previous experiments of the present study. Thus, as in Experiment 3, 66.7 % of entire valid cues remained to be "not so useful" because the cued set of items in VSTM and a single test probe did not contextually match (cue-set-size-3 trials).

The blocked-trial design allowed the observers to anticipate that only a particular cue arrangement would be available for given two blocks in advance (see Awh, Dhaliwal, Christensen, & Matsukura, 2001; Matsukura & Vecera, 2011, for a similar method). That is, in the cue-set-size-1 condition, the observers would expect no contextual mismatch between the cued and probed items. In contrast, in the cue-set-size-3 condition, the observers would expect a discernible contextual mismatch between the cued set of items in VSTM and a single test probe. If memory-level attention is sensitive to the observers' goal and motivation, then the observers should stop relying on multiple retention-interval cues and start performing a simple change-detection task. Consequently, a significant cueing benefit should be observed in the cue-set-size-1 condition but not in the cue-set-size-3 conditions.

Here, it is important to remember that, unlike earlier experiments in the present study, the goal of Experiment 4 was no longer to examine whether attention can select multiple cued items when the cued set of items in VSTM contextually match the to-be-compared percepts in the test array. Because the design of Experiment 4 completely dissociates the observers' mindsets for cue-set-size-1 trials (i.e., encouraged to use the cueing information) and cue-set-size-3 trials (i.e., discouraged to use the cueing information), the question of whether selection of multiple cued items is possible during VSTM maintenance cannot be examined. By replicating Makovski and Jiang's (2007) results in Experiment 2, we have demonstrated that multiple retention-interval cues cannot be efficiently utilized with the single-item recognition task. In Experiment 4, a boundary of the contextual match between the cued and probed items in the cue-set-size-1 condition was examined.

Method

The method was identical to that of Experiment 3 except (1) a new group of 18 observers participated in the experiment, and (2) cue arrangement (single cue, grouped cue, random cue) was blocked. In order to counterbalance the block order, the number of observers was increased from 16 to 18.

While the blocked-trial design can motivate the observers to use the single cue without changing the cue validity of 50 %, overall change-detection accuracy is expected to be higher than that of Experiments 2 and 3 for the following two reasons. First, because the blocked-trial design allows the observers to anticipate ease of single-item cueing trials relative to multiple-item cueing trials, overall accuracy in the cue-set-size-1 condition should be elevated. Consistent with

the previous experiments, the observers in Experiment 4 went through a few minutes of practice trials prior to the experimental session. Thus, each observer would become aware of the contextual mismatch between the cued set of items in VSTM and a single test probe for the multiple-item cueing condition before the experimental session. Based on Matsukura and Vecera (2011) who used a similar blocked-trial design with the feature-report task, we expected roughly 10 % overall accuracy increase.

Second, overall change-detection accuracy in the cue-set-size-3 condition is also expected to be raised because the observers start performing a simple change-detection task by ignoring the cues. A number of retention-interval cueing studies consistently showed that accuracy of neutrally cued (or no-cue) trials was lower than accuracy of typical color change-detection trials without any cueing manipulation. This is despite the fact that a neutrally cued trial in the retention-interval cueing paradigm is conceptually equivalent to a standard change-detection trial. Specifically, the single-item retention-interval cueing task tends to produce approximately 65–70 % of accuracy for neutrally cued trials (memory-array set size 6 with unspeeded responses, e.g., Makovski & Jiang, 2007; Matsukura 2007, 2014). In contrast, a simple color change-detection task tends to yield approximately 75 % of accuracy when the observers are asked to recognize a single test probe at the end of a trial (again for memory-array set size 6 with unspeeded responses, e.g., Exp. 6 in Vogel et al., 2001). Thus, if the observers start ignoring multiple retention-interval cues and performing a simple change-detection task in the cue-set-size-3 condition, then overall accuracy should increase about 10 %.

It is important to note that, given sudden death of VSTM representations cannot be expected until approximately 4 s after the memory array offset (Zhang & Luck, 2009), the longer duration between the memory array offset and the test probe onset in the retention-interval cueing task (relative to the standard 900-ms retention interval of a simple change-detection trial) is unlikely to be a culprit of lower accuracy for neutrally cued trials. Indeed, when we conducted a pilot single-item recognition experiment ($n = 8$) by removing the cue (thus, a typical change-detection trial) but maintaining the identical retention interval (ranging from 2,100 ms to 3,600 ms), accuracy was elevated to 74 %. We acknowledge that, at present, the exact mechanism that produces higher accuracy in canonical change-detection trials than in neutrally cued trials of the retention-interval cueing task is unknown. Although it is not difficult to imagine that experiencing validly cued trials together within a single experiment lowers accuracy of neutrally cued trials, the exact mechanism has not been examined and proposed in the literature yet. Here, elevated accuracy of the standard change-detection task was used as an index that the observers stop relying on the retention-interval cueing information.

Results and discussion

Figure 11 shows mean change-detection accuracy of Experiment 4 as a function of cue type and cue set size. As predicted, while overall accuracy experienced about 10 % increase, the single-item cueing benefit successfully returned.

An ANOVA with within-subjects factors of cue type and cue set size revealed a significant main effect of cue type, $F(1, 17) = 20.77$, $p < .0001$, $\eta_p^2 = .55$, as accuracy was higher in validly cued trials than in neutrally cued trials. Higher accuracy in the cue-set-size-1 condition than in the cue-set-size-3 condition led to a significant main effect of cue set size, $F(1, 17) = 9.51$, $p < .007$, $\eta_p^2 = .36$. The presence of a significant cueing benefit in the cue-set-size-1 condition but not in the cue-set-size-3 condition led to a significant two-way interaction of cue type and cue set size, $F(1, 17) = 13.45$, $p < .002$, $\eta_p^2 = .44$. As apparent in Fig. 11, the planned within-cue-set-size-3 analyses showed neither the main effects of cue type, $F(1, 17) = .00$, $p = .97$, $\eta_p^2 = .00$, perceptual grouping, $F(1, 17) = .03$, $p = .87$, $\eta_p^2 = .00$, nor the interaction of cue type and perceptual grouping, $F(1, 17) = .59$, $p = .45$, $\eta_p^2 = .03$. Planned pair-wise comparisons confirmed that the observed retention-interval cueing benefit was significant in the single-cue condition, $t(17) = 5.33$, $p < .0001$, but neither in the grouped-cue condition, $t(17) = .48$, $p = .64$, nor in the random-cue condition, $t(17) = -.51$, $p = .62$.

The results from Experiment 4 indicate that the disappearance of the single-item cueing benefit in Experiment 3 was driven by the goal-directed selection of memory-level attention. For cue-set-size-1 trials in the single-item recognition experiment, there is no contextual mismatch between the cued and probed items. Despite this fact, when the observers can
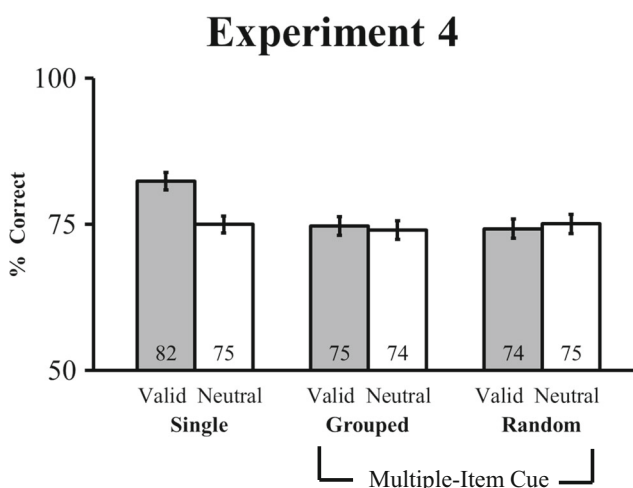

Fig. 11 Mean change-detection accuracy from Experiment 4 (single-item recognition) as a function of cue type (valid, neutral) and cue set size (1 [single] vs. 3 [grouped, random]). Cue arrangement (single, grouped, random) was blocked

figure out that the majority of valid cues are not useful to perform a given change-detection task, they opt out utilizing the retention-interval cue. These results also account for why the single-item cueing benefit was observed only when precue trials were mixed in the experiment (Exp. 2, as well as Makovski & Jiang, 2007).

General discussion

In the present study, we investigated whether attention can select multiple cued items during VSTM maintenance. In line with the sensory recruitment hypothesis, attention can indeed select multiple items represented in VSTM; however, successful selection of multiple cued items requires a reasonable contextual match between the cued representations in VSTM and the to-be-compared percepts. In Experiment 1, we demonstrated that the observers could utilize multiple retention-interval cues when they were asked to recognize a set of the items in the manner that these items had originally been cued (set-item recognition). Moreover, these multiple items can be selected regardless of whether they are perceptually organized into a single group or positioned on noncontiguous locations. However, such a multiple-item cueing benefit disappeared when a discernible contextual mismatch was introduced between the cued set of items and a single test probe (single-item recognition, Exp. 2, which replicated Makovski & Jiang, 2007).

In the process of testing the generality of Makovski and Jiang's (2007) single-item recognition task results, we also replicated and extended the recent finding that memory-level selection is tightly coupled with the observers' goal and motivation (Shimi et al., 2014). Reminiscent to the standard cue validity decrease, when the observers discover that the majority of valid cues are not useful to perform a given change-detection task, they stop utilizing a single retention-interval cue (Exp. 3). Supporting the hypothesis that memory-level attention operates through the goal-directed selection, the single-item cueing benefit returned when the observers were encouraged to use the cueing information (Exp. 4). To our knowledge, this is the first study to demonstrate that a factor other than the cue validity determines whether the observers would utilize the cueing information during the retention interval or not.

Because the present study focused on the question that has been largely unexplored in the current retention-interval cueing literature, we have not yet had a chance to address the following two issues. The first issue pertains to possible selection mechanisms that produced the multiple-item cueing benefits in the set-item recognition experiments. For example, Matsukura et al. (2007) demonstrated that, when attention is directed to an entire hemifield that contains multiple items on contiguous locations, attention serves to protect the cued items

from memory-related degradation processes such as decay, possible interferences by other items stored in VSTM, or some other kind of degradation process that may occur during the retention interval; however, whether or not selection of multiple items on noncontiguous locations operates through such a mechanism remains to be tested in the future. Unfortunately, such an investigation cannot be initiated unless at least one condition that multiple noncontiguous locations can be selected is identified (e.g., set-item recognition), and the present study has accomplished this goal.

The second issue is relevant to possible mechanisms that led to lack of the multiple-item cueing benefit in the single-item recognition experiments. Provided that everything between set-item recognition and single-item recognition tasks is physically identical up until the time of the test, it is tempting to hypothesize that the cued set of items are more likely to be retained in VSTM than other un-cued items for both tasks; though, in the single-item recognition task, the visual system fails to correctly recognize one of the cued items at the time of the test. However, lack of the multiple-item cueing benefit in the single-item recognition task can also be generated through failure to attend to multiple retention-interval cues. Because the observers can clearly attend to multiple cues in the set-item recognition task (Exp. 1), it is unlikely that the visual system fails to attend to multiple retention-interval cues *under any circumstances*. Yet, the observers opt out using multiple retention-interval cues when they know that they will encounter a single test probe at the end of each trial (e.g., Exp. 4). Therefore, if lack of the multiple-item cueing benefit in the single-item recognition experiment is caused by failure to attend to multiple cues, then the possibility that the observers fail to attend to multiple retention-interval cues due to their top-down goal and motivation remains strong. While the current series of experiments are not equipped with differentiating these alternatives, this issue seems ripe for future investigations.

The current series of experiments employed cue set size 3 for the multiple-item cueing condition in order (1) to occupy the typical VSTM capacity of three to four items, (2) to be comparable with the maximal cue set size used in the two groups of studies (Makovski & Jiang, 2007; Matsukura et al., 2007; Williams & Woodman, 2012), and (3) to maximize the contextual discrepancy between the cued set of items in VSTM and a single test probe in the single-item recognition experiments. Within the environment that meets all these three conditions, we found that successful selection of multiple cued items during VSTM maintenance requires a certain level of contextual match between the cued set of items in VSTM and the to-be-compared percepts. Admittedly, the present study did not attempt to identify a variety of the boundary conditions that the contextual discrepancy between the cued memory representations and the to-be-compared percepts can be overcome. However, the findings that memory-level attention is sensitive to the observers'

goal and motivation strongly suggest that it may not be so difficult to overcome such a contextual discrepancy.

For example, Delvenne and Holt (2012) recently reported a significant multiple-item cueing benefit with the single-item recognition task (similar to Exp. 3 in the present study, with all retention-interval cue trials), but only when the cued set of two items was presented across the bilateral hemi-fields. Because the current series of experiments did not examine the effect of stimuli arrangement across two hemifields on the multiple-item cueing benefit, we cannot form any theory regarding the bilateral hemifield cueing advantage per se (it is impossible to split three cued items equally across two hemifields). However, Delvenne and Holt's multiple-item cueing benefit with the single-item recognition task can be sensibly explained by the goal-directed selection mechanism of memory-level attention.

Delvenne and Holt (2012) presented zero, one, and two (bilaterally or unilaterally arranged) retention-interval cues equiprobably with the memory-array set size 8. While the upper limit of cue set size in the present study was three, the upper limit of cue set size in Delvenne and Holt's study was two, which is well below the typical VSTM capacity. Thus, the comparison between the cued set of two items in VSTM and a single test probe in perception is not likely to suffer from a huge contextual discrepancy. Having only cue-set-size-1 and -2 trials in the experiment might have also motivated the observers to utilize the cueing information. As shown in the present study, if the observers in Delvenne and Holt evaluated that use of one and two retention-interval cues was easy and beneficial enough to perform a given change-detection task, then it is plausible to observe the multiple-item cueing benefit. In this regard, the observers in Makovski and Jiang (2007, Exp. 2A) who encountered zero (i.e., equivalent to six), one, two, and three cues might have been discouraged to utilize two retention-interval cues, as the upper limit of cue set size in their experiment was three.

Curiously though, Makovski and Jiang (2007, Exp. 2B), which is a different experiment from their Exp. 2A) reported the opposite pattern of results from Delvenne and Holt (2012). That is, when Makovski and Jiang conducted the single-item recognition task with retention-interval cue trials alone but with the central arrow cue, they failed to observe any multiple-item cueing benefit. Like Delvenne and Holt, Makovski and Jiang presented zero, one, and two retention-interval cues equiprobably but with the memory-array set size 6. Here, one interesting characteristic of this experiment was that, for cue-set-size-2 trials, the cue was a straight bidirectional arrow that *always* pointed to two *diagonal* locations represented in VSTM. Because Makovski and Jiang used memory-array set size 6 with a horizontal axis, this "diagonal" restriction left only two possible directions that the straight bidirectional arrow could point at the same time (i.e., no

horizontal split). This means that, for cue-set-size-2 trials, the observers' attention was always directed to two items presented across the bilateral hemifields. Yet Makovski and Jiang failed to observe any multiple-item cueing benefit. For unknown reasons, Delvenne and Holt did not discuss this lack of the bilateral hemifield cueing benefit reported by Makovski and Jiang. While identifying the exact source of the contradicting results that emerged between these two studies is beyond the present study's scope, the origin of this difference should be determined in future studies.

Of course, ways to reduce the contextual mismatch between the cued memory representations and the to-be-compared percepts should not be limited to shrinking the distance between cue set size (how many items in VSTM are cued) and the number of probes. We expect that the weights among stimulus type, memory-array set size, cue set size, and the number of probes play some roles in determining a reasonable contextual match that makes selection of multiple cued items during VSTM maintenance possible. For example, a recent magneto-encephalography study (Poch, Campo, & Barnes, 2014), which adopted the single-item recognition task, reported a statistical trend for the multiple-item cueing benefit with cue set size 2, memory-array set size 4 and orientation bar stimuli. However, lack of the whole statistical report as well as the behavioral task demand details (e.g., unspeeded or speeded responses) require us to carefully interpret these results. Before embarking on systematic investigations of the boundary conditions that the contextual discrepancy between the cued memory representations and the to-be-compared percepts can be overcome, the present study first demonstrated that the visual system cannot survive the apparent contextual discrepancy between the cued set of three items in VSTM and a single test probe.

In conclusion, although the findings from the present study raise many interesting questions regarding selective attention mechanisms that operate during VSTM maintenance, it is clear that (1) attention can select more than a single item during VSTM maintenance when the observers are asked to recognize a set of items in the manner that these items were originally cued, and (2) attention can select multiple cued items regardless of whether these items are perceptually organized into a single group (contiguous locations) or not (noncontiguous locations).

Appendix

Table 1 ANOVAs based on mean d' values (Macmillan & Creelman, 1991)

	df	F	p	η_p^2
Experiment 1A				
Main effect: Cue type	1, 31	31.95	<.0001	.51
Main effect: Cue set size	1, 31	32.67	<.0001	.51
Interaction: Cue type × Cue set size	1, 31	.10	.74	.00
Main effect: Cue type	1, 31	21.32	<.0001	.41
Main effect: Perceptual grouping	1, 31	8.44	.007	.21
Interaction: Cue type × Perceptual grouping	1, 31	.99	.33	.03
Experiment 1B				
Main effect: Cue type	1, 15	33.02	<.0001	.60
Main effect: Cue set size	1, 15	22.89	<.0001	.69
Interaction: Cue type × Cue set size	1, 15	5.36	.04	.26
Main effect: Cue type	1, 15	10.71	.005	.42
Main effect: Perceptual grouping	1, 15	12.12	.003	.45
Interaction: Cue type × Perceptual grouping	1, 15	.07	.79	.01
Experiment 2				
Main effect: Precue type	1, 31	143.75	<.0001	.82
Main effect: Cue set size	1, 31	65.67	<.0001	.68
Interaction: Precue type × Cue set size	1, 31	60.20	<.0001	.66
Main effect: Precue type	1, 31	54.00	<.0001	.64
Main effect: Perceptual grouping	1, 31	1.55	.22	.05
Interaction: Precue type × Cue set size	1, 31	6.44	.02	.17
Main effect: Retention-interval cue type	1, 31	14.93	.001	.33
Main effect: Cue set size	1, 31	1.65	.21	.50
Interaction: Retention-interval cue type × Cue set size	1, 31	9.58	.004	.24
Main effect: Retention-interval cue type	1, 31	.47	.50	.02
Main effect: Perceptual grouping	1, 31	.02	.88	.00
Interaction: Retention-interval cue type × Perceptual grouping	1, 31	.14	.71	.01
Interaction: Cue presentation timing × Cue set size	1, 31	41.60	<.0001	.57
Experiment 3				
Main effect: Cue type	1, 15	.23	.88	.00
Main effect: Cue set size	1, 15	3.61	.08	.19
Interaction: Cue type × Cue set size	1, 15	3.46	.83	.19
Experiment 4				
Main effect: Cue type	1, 17	10.77	.0004	.39
Main effect: Cue set size	1, 17	4.67	.05	.22
Interaction: Cue type × Cue set size	1, 17	13.07	.002	.44
Main effect: Cue type	1, 17	.29	.78	.01
Main effect: Perceptual grouping	1, 17	.08	.60	.02
Interaction: Cue type × Perceptual grouping	1, 17	.79	.39	.04

Table 2 Planned pairwise comparisons based on mean d' values

	Valid d'	Neutral d'	df	t	p
Experiment 1A					
Single cue trials	1.3	1.0	31	3.34	.002
Grouped cue trials	1.0	.8	31	2.81	.009
Random cue trials	.9	.6	31	3.76	.0007
Experiment 1B					
Single cue trials	1.7	1.2	15	5.08	<.0001
Grouped cue trials	1.3	1.0	15	2.28	.037
Random cue trials	1.0	.8	15	2.55	.02
Experiment 2					
Single precue trials	3.3	1.5	31	12.33	<.0001
Grouped precue trials	2.4	1.5	31	6.84	<.0001
Random precue trials	2.1	1.6	31	4.14	.0002
Single retention-interval cue trials	1.3	.8	31	3.97	.0004
Grouped retention-interval cue trials	.9	.9	31	.12	.91
Random retention-interval cue trials	1.0	.9	31	.91	.37
Experiment 4					
Single cue trials	2.1	1.6	17	4.44	.0003
Grouped cue trials	1.6	1.5	17	.22	.83
Random cue trials	1.5	1.6	17	1.00	.33

Table 3 Mean d' values for two types of valid random-cue trials in the set-item recognition task (Exp. 1)

	Fig. 4a	Fig. 4b	df	t	p
Experiment 1A					
Valid random cue trials	.9	.9	31	.06	.95
Experiment 1B					
Valid random cue trials	1.0	.9	15	.51	.61

Figure 4a represents a trial that each of three dot cues appeared at every other item's location. Figure 4b represents a trial that two out of three cues appeared at adjacent locations while a single cue appeared in isolation.

References

Albers, A. M., Kok, P., Toni, I., Dijkerman, H. C., & de Lange, F. P. (2013). Shared representations for working memory and mental imagery in early visual cortex. *Current Biology, 23,* 1427–1431. doi:10.1016/j.cub.2013.05.065

Alvarez, G. A., Gill, J., & Cavanagh, P. (2012). Anatomical constraints on attention: Hemifield independence is a signature of multifocal spatial selection. *Journal of Vision, 12*(5):9, 1–20. doi:10.1167/12.5.9.

Anderson, D. E., Ester, E. E., Serences, J. T., & Awh, E. (2013). Attending multiple items decreases the selectivity of population responses in human primary visual cortex. *Journal of Neuroscience, 33,* 9273–9282. doi:10.1523/JNEUROSCI.0239-13.2013

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

Averbach, E., & Coriell, A. S. (1961). Short-term memory in vision. *Bell System Technical Journal, 40,* 309–328. http://onlinelibrary.wiley.com/doi/10.1002/j.1538-7305.1961.tb03987.x/abstract

Awh, E., Dhaliwal, H., Christensen, S., & Matsukura, M. (2001). Evidence for two components of object-based selection. *Psychological Science, 12,* 329–334. doi:10.1111/1467-9280.00360

Awh, E., & Jonides, J. (2001). Overlapping mechanisms of attention and spatial working memory. *Trends in Cognitive Sciences, 5,* 119–126. doi:10.1016/S1364-6613(00)01593-X

Awh, E., & Pashler, H. (2000). Evidence for split attentional foci. *Journal of Experimental Psychology: Human Perception and Performance, 26,* 834–846. doi:10.1037/0096-1523.26.2.834

Baddeley, A. D. (1986). *Working memory.* Oxford, UK: Oxford University Press, Clarendon Press.

Berryhill, M. E., Richmond, L. L., Shay, C. S., & Olson, I. R. (2012). Shifting attention among working memory representations: Testing cue type, awareness, and strategic control. *Quarterly Journal of Experimental Psychology, 65,* 426–438. doi:10.1080/17470218.2011.604786

Besner, D., Davies, J., & Daniels, S. (1981). Reading for meaning: The effects of concurrent articulation. *Quarterly Journal of Experimental Psychology, 33A,* 415–437. doi:10.1080/14640748108400801

Bundesen, C. (1990). A theory of visual attention. *Psychological Review, 97,* 523–547. doi:10.1037/0033-295X.97.4.523

Christophel, T. B., Hebart, M. N., & Haynes, J.-D. (2012). Decoding the contents of visual short-term memory from human visual and parietal cortex. *Journal of Neuroscience, 32,* 12983–12989. doi:10.1523/JNEUROSCI.0184-12.2012

Delvenne, J.-F., & Holt, J. L. (2012). Splitting attention across the two visual fields in short-term memory. *Cognition, 122,* 258–263. doi:10.1016/j.cognition.2011.10.015

Egly, R., Driver, J., & Rafal, R. D. (1994). Shifting visual attention between objects and locations: Evidence from normal and parietal lesion subjects. *Journal of Experimental Psychology: General, 123,* 161–177. doi:10.1037/0096-3445.123.2.161

Ester, E. F., Anderson, D. E., Serences, J. T., & Awh, E. (2013). A neural measure of precision in visual working memory. *Journal of Cognitive Neuroscience, 25,* 754–761. doi:10.1162/jocn_a_00357

Ester, E. F., Fukuda, K., May, L. M., Vogel, E. K., & Awh, E. (2014). Evidence for a fixed capacity limit in attending multiple locations. *Cognitive, Affective, & Behavioral Neuroscience, 14,* 62–77. doi:10.3758/s13415-013-0222-2

Franconeri, S. L., Alvarez, G. A., & Enns, J. T. (2007). How many locations can be selected at once? *Journal of Experimental Psychology: Human Perception and Performance, 33,* 1003–1012. doi:10.1037/0096-1523.33.5.1003

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

Hahn, S., & Kramer, A. F. (1998). Further evidence for the division of attention between noncontiguous locations. *Visual Cognition, 5,* 217–256. doi:10.1080/713756781

Harrison, S. A., & Tong, F. (2009). Decoding reveals the contents of visual working memory in early visual areas. *Nature, 458,* 632–635. doi:10.1038/nature07832

Hecht, L. N., & Vecera, S. P. (2007). Attentional selection of complex objects: Joint effects of surface uniformity and part structure. *Psychonomic Bulletin & Review, 14,* 1205–1211. doi:10.3758/BF03193114

Hollingworth, A., & Hwang, S. (2013). The relation between visual working memory and attention: Retention of precise color information in the absence of effects on perceptual selection. *Philosophical Transactions of the Royal Society B: Biological Sciences, 368,* 1–9. doi:10.1098/rstb.2013.0061

Hollingworth, A., & Maxcey-Richard, A. M. (2013). Selective maintenance in visual working memory does not require sustained visual attention. *Journal of Experimental Psychology: Human Perception and Performance, 39,* 1047–1058. doi:10.1037/a0030238

Irwin, D. E., & Yeomans, J. M. (1986). Sensory registration and informational persistence. *Journal of Experimental Psychology: Human Perception and Performance, 12,* 343–360. doi:10.1037/0096-1523.12.3.343

Janczyk, M., & Berryhill, M. E. (2014). Orienting attention in visual working memory requires central capacity: Decreased retro-cue effects under dual-task conditions. *Attention, Perception, & Psychophysics, 76,* 715–724. doi:10.3758/s13414-013-0615-x

Jonides, J. (1981). Voluntary versus automatic control over the mind's eye's movement. In J. Long & A. Baddeley (Eds.), *Attention and performance IX* (pp. 187–203). Hillsdale, NJ: Erlbaum.

Kramer, A. F., & Hahn, S. (1995). Splitting the beam: Distribution of attention over noncontiguous regions of the visual field. *Psychological Science, 6,* 381–386. doi:10.1111/j.1467-9280.1995.tb00530.x

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

Loftus, G. R., & Masson, M. E. J. (1994). Using confidence intervals in within-subject designs. *Psychonomic Bulletin & Review, 1,* 476–490. doi:10.3758/BF03210951

Luck, S. J., & Vecera, S. P. (2002). Attention. In S. Yantis (Ed.), *Stevens' handbook of experimental psychology: Vol. 1: Sensation and perception* (Vol. 3, pp. 235–286). New York: Wiley.

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

Luck, S. J., & Vogel, E. K. (2013). Visual working memory capacity: From psychophysics and neurobiology to individual differences. *Trends in Cognitive Sciences, 17,* 391–400. doi:10.1016/j.tics.2013.06.006

Macmillan, N. A., & Creelman, C. D. (1991). *Detection theory: A user's guide.* Cambridge, UK: Cambridge University Press.

Makovski, T. (2012). Are multiple visual short-term memory storages necessary to explain the retro-cue effect? *Psychonomic Bulletin & Review, 19,* 470–476. doi:10.3758/s13423-012-0235-9

Makovski, T., & Jiang, Y. V. (2007). Distributing versus focusing attention in visual short-term memory. *Psychonomic Bulletin & Review, 14,* 1072–1078. doi:10.3758/BF03193093

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., Cosman, J. D., Roper, Z. J. J., Vatterott, D. B., & Vecera, S. P. (2014). Location-specific effects of attention during visual short-term memory maintenance. *Journal of Experimental Psychology: Human Perception and Performance, 40,* 1103–1116. doi:10.1037/a0035685

Matsukura, M., & Hollingworth, A. (2011). Does visual short-term memory have a high-capacity stage? *Psychonomic Bulletin & Review, 18,* 1098–1104. doi:10.3758/s13423-011-0153-2

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

Matsukura, M., & Vecera, S. P. (2006). The return of object-based attention: Selection of multiple-region objects. *Perception & Psychophysics, 68,* 1163–1175. doi:10.3758/BF03193718

Matsukura, M., & Vecera, S. P. (2011). Object-based selection from spatially-invariant representations: Evidence from a feature-report task. *Attention, Perception, & Psychophysics, 73,* 447–457. doi:10.3758/s13414-010-0039-9

Mazyar, H., van den Berg, R., & Ma, W. J. (2012). Does precision decrease with set size? *Journal of Vision, 12*(6):10, 1–16. doi:10.1167/12.6.10

Munneke, J., Belopolsky, A. V., & Theeuwes, J. (2012). Shifting attention within memory representations involves early visual areas. *PloS One, 7,* e35528. doi:10.1371/journal.pone.0035528

Murray, D. J. (1968). Articulation and acoustic confusability in short-term memory. *Journal of Experimental Psychology, 78,* 679–684. doi:10.1037/h0026641

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

Palmer, J. (1990). Attentional limits on the perception and memory of visual information. *Journal of Experimental Psychology: Human Perception and Performance, 16,* 332–350. doi:10.1037/0096-1523.16.2.332

Palmer, E. M., Fencsik, D. E., Flusberg, S. J., Horowitz, T. S., & Wolfe, J. M. (2011). Signal detection evidence for limited capacity in visual search. *Attention, Perception, & Psychophysics, 73,* 2413–2424. doi:10.3758/s13414-011-0199-2

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

Poch, C., Campo, P., & Barnes, G. R. (2014). Modulation of alpha and gamma oscillations related to retrospectively orienting attention within working memory. *European Journal of Neuroscience, 40,* 2399–2405. doi:10.1111/ejn.12589

Posner, M. I., & Cohen, Y. (1984). Components of visual orienting. In H. Bouma & D. G. Bouwhuis (Eds.), *Attention and performance X: Control of language processes* (pp. 531–556). Hillsdale, NJ: Erlbaum.

Posner, M. I., Snyder, C. R., & Davidson, B. J. (1980). Attention and the detection of signals. *Journal of Experimental Psychology: General, 109,* 160–174. doi:10.1037/0096-3445.109.2.160

Pratte, M. S., & Tong, F. (2014). Spatial specificity of working memory representations in the early visual cortex. *Journal of Vision, 14*(3): 22, 1–12. doi:10.1167/14.3.22

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

Rerko, L., Souza, A. S., & Oberauer, K. (2014). Retro-cue benefits in working memory without sustained focal attention. *Memory & Cognition, 42,* 712–728. doi:10.3758/s13421-013-0392-8

Rouder, J. N., Morey, R. D., Morey, C. C., & Cowan, N. (2011). How to measure working memory capacity in the change detection paradigm. *Psychonomic Bulletin & Review, 18,* 324–330. doi:10.3758/s13423-011-0055-3

Schmidt, B. K., Vogel, E. K., Woodman, G. F., & Luck, S. J. (2002). Voluntary and automatic attentional control of visual working memory. *Perception & Psychophysics, 64,* 754–763. doi:10.3758/BF03194742

Serences, J. T., Ester, E. F., Vogel, E. K., & Awh, E. (2009). Stimulus-specific delay activity in human primary visual cortex. *Psychological Science, 20,* 207–214. doi:10.1111/j.1467-9280.2009.02276.x

Shimi, A., Nobre, A. C., Astle, D., & Scerif, G. (2014). Orienting attention within visual short-term memory: Development and mechanisms. *Child Development, 85,* 578–592. doi:10.1111/cdev.12150

Sligte, I. G., Scholte, H. S., & Lamme, V. A. F. (2008). Are there multiple visual short-term memory stores? *PloS One, 3,* e1699. doi:10.1371/journal.pone.0001699

Souza, A. S., Rerko, L., Lin, H.-Y., & Oberauer, K. (2014). Focused attention improves working memory: Implications for flexible-resource and discrete-capacity models. *Attention, Perception, & Psychophysics, 76,* 2080–2102. doi:10.3758/s13414-014-0687-2

Souza, A. S., Rerko, L., & Oberauer, K. (2014). Unloading and reloading working memory: Attending to one item frees capacity. *Journal of Experimental Psychology: Human Perception and Performance, 40,* 1237–1256. doi:10.1037/a0036331

Sperling, G. (1960). The information available in brief visual presentations. *Psychological Monograph: General and Applied, 74*(11, Whole No. 498), 1–29. doi:10.1037/h0093759

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

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

Williams, M., & Woodman, G. F. (2012). Directed forgetting and directed remembering in visual working memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 38,* 1206–1220. doi:10.1037/a0027389

Wojciulik, E., & Kanwisher, N. (1999). The generality of parietal involvement in visual attention. *Neuron, 23,* 747–764. doi:10.1016/S0896-6273(01)80033-7

Woodman, G. F., Vecera, S. P., & Luck, S. J. (2003). Perceptual organization influences visual working memory. *Psychonomic Bulletin & Review, 10,* 80–87. doi:10.3758/BF03196470

Yantis, S. (2000). Goal-directed and stimulus-driven determinants of attentional control. In S. Monsell & J. Driver (Eds.), *Control of cognitive processes: Attention and performance XVIII* (pp. 73–103). Cambridge, MA: MIT Press.

Zhang, W., & Luck, S. J. (2009). Sudden death and gradual decay in visual working memory. *Psychological Science, 20,* 423–428. doi:10.1111/j.1467-9280.2009.02322.x