Estimated capacity of object files in visual short-term memory is not improved by retrieval cueing

Jun Saiki  
Graduate School of Human and Environmental Studies,  
Kyoto University, Japan

Hirofumi Miyatsuji  
Graduate School of Informatics, Kyoto University, Japan

Visual short-term memory (VSTM) has been claimed to maintain three to five feature-bound object representations. Some results showing smaller capacity estimates for feature binding memory have been interpreted as the effects of interference in memory retrieval. However, change-detection tasks may not properly evaluate complex feature-bound representations such as triple conjunctions in VSTM. To understand the general type of feature-bound object representation, evaluation of triple conjunctions is critical. To test whether interference occurs in memory retrieval for complete object file representations in a VSTM task, we cued retrieval in novel paradigms that directly evaluate the memory for triple conjunctions, in comparison with a simple change-detection task. In our multiple object permanence tracking displays, observers monitored for a switch in feature combination between objects during an occlusion period, and we found that a retrieval cue provided no benefit with the triple conjunction tasks, but significant facilitation with the change-detection task, suggesting that low capacity estimates of object file memory in VSTM reflect a limit on maintenance, not retrieval.

Keywords: Visual short-term memory, feature binding, change detection


Introduction

Our visual world contains numerous objects, each with multiple features. To perceive the world properly, the brain must represent which features belong to each object. This process of feature integration has often been discussed under the term “binding problem,” which has primarily been studied in the context of visual perception. Treisman and Schmidt (1982) found evidence suggesting that feature binding in visual perception requires visual attention. How, then, are object representations with integrated features stored in visual short-term memory (VSTM)? At least two possibilities present themselves:

1. an attentional mechanism may form stable representations of objects, which can be maintained in VSTM; or
2. feature-bound object representations may not necessarily be maintained within VSTM.

This basic problem of feature binding in visual memory remains unresolved.

Luck and Vogel (1997) showed that when performing a short-term change-detection task with multifeature objects, humans have a capacity of about four objects. Although this finding of a capacity limit of three to five objects in VSTM has been replicated repeatedly (Todd & Marois, 2004; Vogel & Machizawa, 2004), the issue of the unit of VSTM as feature-bound objects remains equivocal. One problem with the original study was that the stimulus design is unsuitable for evaluating feature binding. As with the stimuli used in perceptual feature binding, manipulation of the combination of features must be performed while keeping the identities of component features constant. Some recent studies (Saiki, 2003a, 2003b; Wheeler & Treisman, 2002) used modified tasks to satisfy this requirement, revealing that estimated capacity is significantly smaller than three to five objects. In these studies, the change to be detected is sometime a switching of features between two objects (called binding condition), so that the pairing of features changes, but the feature identities do not. The present study investigated the nature of limitations on feature-bound memory in VSTM by examining whether previous findings of highly limited capacity using single feature conjunctions can be generalized to more complex situations.

Manipulations to investigate the nature of binding memory

To understand performance in measures of short-term memory binding, both memory representation and cognitive processing in the task should be considered. Different kinds of memory representations have been postulated for VSTM. Feature-bound memory representation (Luck & Vogel, 1997), unbound memory for features (Wheeler & Treisman, 2002), memory for configuration (Jiang, Olson, & Chun, 2000), and possibly others may be components...
of VSTM. Memory for feature binding has often been discussed under the notion of object files (Kahneman, Treisman, & Gibbs, 1992), so this study uses the term “object files” to refer to memory for feature bindings in general. In this work, we distinguish two types of object files. The first is a complete object file, representing the complete set of features of an object. For example, if an object is presented with different shapes and colors, the complete object file comprises a conjunction of shape, color, and location. The second is a partial object file, representing a partial set of features. Using the example above, a partial object file would comprise a conjunction of color and location, shape and location, or color and shape. The present study utilized objects defined by a combination of location, shape, and color, so a complete object file corresponds to memory for triple conjunctions (location, shape, and color), and partial object files correspond to memory for single conjunctions.

Considering the cognitive processing involved in the task, performance may reflect encoding, maintenance, retrieval, and/or comparison. A deficit in a memory task may thus be caused by a limit in storage capacity or by a bottleneck in memory retrieval and/or comparison between memory and perceptual representations. In the present study, memory retrieval and memory comparison could not be experimentally distinguished, so we have used the term memory retrieval to refer to both in the following.

Some studies have suggested that low estimated capacity for feature binding memory relative to feature memory may reflect differences in memory retrieval. Wheeler and Treisman (2002) compared the single-probe paradigm, where only one object was presented in the probe display to be judged for the presence of change, with the multiple-probe paradigm, where the whole probe display needed to be compared with the initial display. They showed that the single-probe condition significantly improved performance in the binding condition compared to the multiple-probe condition. This improvement in task performance can be interpreted as a reduction of interference and/or a facilitation of memory retrieval by the single probe. Allen, Baddeley, and Hitch (2006) reported similar results.

The single probe advantage in the binding condition leaves some questions open regarding the nature of representation and processing in VSTM. First, the findings of Wheeler and Treisman (2002) do not necessarily imply that memory for object files in general suffer from a retrieval bottleneck in the multiple probe condition. Wheeler and Treisman investigated complete object files from color–location conjunction stimuli and partial object files (shape and color) from triple conjunction stimuli, so whether a single probe advantage is observed with complete object files from triple conjunction stimuli remains unknown. If the previous findings reflect the nature of object files in general, the single probe advantage should be observed with complete object files from triple conjunction stimuli. Conversely, if the previous findings hold true only in certain special situations, the single probe advantage may be limited to partial object files from triple conjunction stimuli.

To address this issue, however, the simple change-detection task with triple conjunction stimuli is insufficient because triple conjunction representations are not necessary to simply detect a change. For example, suppose objects with color and shape are represented as independent sets of two partial object files: color–location and shape–location. This representational scheme is sufficient to detect a change with triple conjunction stimuli by monitoring a change in two sets of partial object files independently. To deal with this problem, Saiki and Miyatsuji (2007) devised a type-identification paradigm. Following the logic of perceptual feature binding studies (Treisman & Schmidt, 1982), the type-identification paradigm requires information on triple feature combination, not just simple conjunctions. Unlike simple change detection, type identification requires the discrimination of sources of change by asking participants to identify a type of switch event. This forces participants to take into account the triple conjunction of shape, color, and location. This task thus allows evaluation of memory for more complex feature representations. Saiki and Miyatsuji applied this task to an experimental paradigm called multiple object permanence tracking (MOPT), which is similar to the binding conditions described by Wheeler and Treisman (2002) in logic, but is also able to evaluate any effect of object motion (Figure 1). A series of experiments revealed that (1) even when objects are stationary, task performance was quite poor; and (2) object motion further impaired performance, even if motion speed was slow. Memory capacity estimated using a standard formula with a proper modification (see Appendix A) was only about 1.5 objects when stationary and 1 object when moving. Earlier studies (Saiki, 2003a, 2003b) showed that this impairment by object motion is not simply due to failure in object tracking but reflects an additional cost of spatiotemporal updating of feature binding. MOPT is suitable to investigate spatiotemporal updating, which is an important characteristic of object file representation.

The type-identification paradigm used in the present study can evaluate memory for object files more properly, but whether deficits in task performance reflect retrieval or maintenance per se remains elusive. Type-identification performance may be impaired not because object files have smaller maintenance capacity than other types of VSTM, but because memory retrieval is more difficult for object files than for unbound features, as suggested by Wheeler and Treisman (2002).

**Research strategy: Combined manipulations**

Taken together, although the manipulations discussed above, of single probe condition and type-identification task are certainly beneficial to elucidate the nature of binding memory, each alone is insufficient. The critical
problem is to evaluate whether performance impairment observed in an experiment reflects memory retrieval or maintenance. The single probe condition used by Wheeler and Treisman (2002) directly manipulated retrieval but left the nature of the memory representation ambiguous. In contrast, the type-identification task can specifically evaluate memory for complete object files, but the task performance may reflect either memory maintenance or retrieval. Therefore, in the present study, retrieval cueing and memory task manipulation were combined to achieve a better understanding of the nature of binding memory.

Experiment 1 combined the type-identification task and retrieval cuing using the MOPT paradigm. A cue indicating the changing object was 100% valid and was presented either just before (precue) or after (postcue) a change occurred. If a cue is effective, the precue condition is expected to show significantly better task performance compared with the no-cue control. The critical condition was the postcue condition. Estimated capacity from behavioral data is determined by two factors: maintenance capacity and costs in memory retrieval. Because the postcue condition substantially reduces retrieval costs, estimated capacity in the postcue condition is closer to the genuine maintenance capacity than in the no-cue condition. The performance facilitation by the postcue condition thus suggests that estimated capacity in the no-cue condition suffers from retrieval costs, while the lack thereof suggests that the estimated capacity in the no-cue condition reflects genuine maintenance capacity. If the difficulty in MOPT observed in Saiki and Miyatsuji (2007) reflects memory retrieval and the single probe benefit observed in Wheeler and Treisman (2002) reflects the nature of complete object files, the postcue will facilitate performance. Also, using the moving condition of MOPT, we compared effects of retrieval cuing between spatiotemporal updating and simple maintenance of complete object files. As shown below, Experiment 1 failed to obtain facilitation by the retrieval cue, suggesting that the retrieval cue benefit occurs only for partial object files, but not for complete object files. Subsequent experiments further examined some alternative interpretations, and the overall results support the notion that retrieval cues improve estimated capacity for partial object files, not for complete object files.

**Experiment 1**

**Method**

**Participants**

Six graduate students at Kyoto University, including one author (HM), participated in the experiment, and all displayed normal color vision.

**Materials**

Participants were shown a pattern of four colored objects with an occluder on top. Smooth rotation of the pattern and occluder at constant angular velocities resulted in alternating appearance and disappearance of the pattern (Movies 1 and 2).

The four colored objects were configured in a diamond pattern, with each object placed at a visual angle of 2.9° from the center of the occluder. Objects were colored using four equiluminant colors (20.85 cd/m², red [CIE \(x = 0.56, y = 0.34\)]; green [\(x = 0.28, y = 0.60\)]; blue [\(x = 0.19, y = 0.14\)]; and yellow [\(x = 0.43, y = 0.49\)]), and combinations of these colors were counterbalanced across trials. Shapes used for objects in the experiment were circle, square, hexagon, and triangle. The colored pattern was occluded using a gray windmill-shaped occluder.
between the two possibilities. The flashing cue was presented for 47 ms at the middle of a visible period of 518 ms. A cue was thus presented 259 ms after an object appeared. The postcue was the same as the precue but was presented after the switch. The no-cue condition was the control condition, with no cue presented (Figure 3). Experimental programs were written in MATLAB, using Psychophysics Toolbox extensions (Brainard, 1997; Pelli, 1997).

**Procedure**

Each experimental trial began with a key press by a participant. After the beep, the initial display with objects and occluder were stationary for 500 ms. The moving sequence then began, followed by the appearance of three response boxes for event types (color, shape, and color-and-shape). The participant selected responses by clicking a response box. To avoid verbal encoding of color and shape, articulatory suppression was used by getting subjects to say “da, da, da.” The entire experiment comprised three experimental sessions, each containing 216 trials. Participants performed one session a day. Within each session, cueing and object motion conditions were randomly mixed from trial to trial. Under each cueing condition, each object motion condition comprised 108 trials, with 36 trials for each event type, resulting in a total of 648 experimental trials.

**Results and discussion**

Figure 4 shows the proportions of correct type identification as a function of object motion for three
cueing conditions. First, the effect of precue was evaluated by comparison with the no-cue condition. A strong effect of precue was seen under both stationary and moving conditions. Analysis of variance (ANOVA) with a 2 (cueing: precue and no-cue) × 2 (object motion: stationary and moving) design revealed significant main effects of cueing ($F(1,5) = 139.01, p < .001$) and object motion ($F(1,5) = 71.78, p < .001$) and their interaction ($F(1,5) = 7.05, p < .05$). The interaction reflects that object motion impaired performance in the no-cue condition, but not in the precue condition. The lack of motion effect in the precue condition possibly reflects the ceiling effect. Thus, if attention can be focused in advance, a feature-bound representation can be maintained for the attended object, regardless of motion. Second, no effect of postcue was seen at all (Figure 4). ANOVA with a 2 (cueing: postcue and no-cue) × 2 (object motion: stationary and moving) design revealed the significant main effect of object motion ($F(1,5) = 27.27, p < .01$). Stationary motion ($F(1,5) = 71.78, p < .001$) and their interaction ($F(1,5) = 7.05, p < .05$). The interaction reflects that object motion impaired performance in the no-cue condition, but not in the precue condition. The lack of motion effect in the precue condition possibly reflects the ceiling effect. Thus, if attention can be focused in advance, a feature-bound representation can be maintained for the attended object, regardless of motion. Second, no effect of postcue was seen at all (Figure 4). ANOVA with a 2 (cueing: postcue and no-cue) × 2 (object motion: stationary and moving) design revealed the significant main effect of object motion ($F(1,5) = 27.27, p < .01$). Stationary
condition showed better performance. No advantage of postcue was observed ($F(1,5) = 1.84, p > .1$). Object motion impaired performance, which is consistent with previous work using MOPT paradigm (Saiki, 2003b; Saiki & Miyatsuji, 2007).

Different event types (color, shape, and both) may have different effects of postcue and object motion. Proportion correct data were next decomposed into event types, and ANOVA with a $2$ (cueing: postcue and no-cue) $\times 2$ (object motion: stationary and moving) $\times 3$ (event type: color, shape, and both) was conducted. In addition to a significant main effect of object motion, a significant main effect of event type ($F(2,10) = 9.16, p < .01$) was identified. None of the event types showed significant main effect of cue, nor interaction with cue, suggesting that different event types did not significantly modulate the effect of postcue. Furthermore, to eliminate a possibility that the cueing effects were concealed by response bias, we decomposed contingency of event types and responses into similarity and response bias factors based on a similarity choice model (Luce, 1959; Townsend & Landon, 1982). Table 1 shows estimated parameters and $\chi^2$ statistic for goodness of fit test for each category. The model fit was reasonably good (model was not rejected at $p = .05$ level), and the values of similarity parameters that reflect confusability are consistent with accuracy data, such as a substantial effect of object motion and no effect of cueing, suggesting that response bias was not responsible for the lack of cueing effects. Taken together, while precue displayed quite strong facilitatory effects on type-identification performance regardless of object motion, postcue had no effects, even when each event type was separately analyzed. However, the lack of postcue effects for each event type may have been due to insufficient statistical power.

As Experiment 1 mixed precue and postcue in the same session, and precue was obviously quite effective, participants may have intentionally focused on precues and ignored postcues, despite being instructed that two kinds of cues were included. To test this possibility, an additional experiment was conducted where the precue condition was eliminated. Results were the same, revealing that only object motion displayed a significant main effect ($F(1,5) = 8.92, p < .05$). Clearly, the lack of postcue effect was not due to strategic ignorance by inclusion of precue. Even when participants knew that all cues were for memory retrieval, postcue could not be utilized effectively.

The lack of postcue effects in this experiment suggests that the deficit for object files observed in MOPT primarily reflects memory maintenance, and that single probe improvement in Wheeler and Treisman (2002) may be specific to partial object files. One alternative account, however, is that the complexity of the type-identification paradigm eliminated any postcue benefit. Response mapping in type identification is more complex than simple change detection. Postcue might indeed have facilitated retrieval of memory for changing object, but comparison of perceptual and memorial representations and response selection in the type identification was so complex that attentional resources to maintain object files were redirected to the mechanism of comparison. To examine this possibility, the next experiment used a task with two response alternatives, as in a typical change-detection task.

### Experiment 2

In Experiment 2, a relevant-feature switch detection task was used. In each trial, the participant was instructed to monitor either color or shape and required to judge whether the stimulus sequence included a switch event on the presupposed feature dimension. This task had only two response alternatives, as in a simple change-detection task, but required distinction between color- and shape-switch events. In principle, these two events can be distinguished in two ways. First, comparison of two complete object files of shape, color, and location leads to correct judgments as in the type-identification task. Alternatively, by selectively encoding and maintaining a partial object file representing the task-relevant feature pair, comparison of a partial object file in memory with perceptual input leads to correct judgment. Note that these two ways differ from nonselective use of partial object files that is sufficient for simple change detection. Thus, if the single-probe advantage found by Wheeler and Treisman (2002) also occurs for complete object files, or for selectively maintained partial object files, facilitation by postcue will be expected.

<table>
<thead>
<tr>
<th>Similarity</th>
<th>Bias</th>
<th>Goodness of fit</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Color</td>
<td>Shape</td>
</tr>
<tr>
<td>Color–shape</td>
<td>0.27</td>
<td>0.26</td>
</tr>
<tr>
<td>Color–both</td>
<td>0.26</td>
<td>0.26</td>
</tr>
<tr>
<td>Shape–both</td>
<td>0.38</td>
<td>0.35</td>
</tr>
<tr>
<td></td>
<td>0.36</td>
<td>0.35</td>
</tr>
</tbody>
</table>

Table 1. Parameter estimates and $\chi^2$ value for goodness of fit test for contingency analysis between event types and responses.
Method

The method was the same as in Experiment 1, except for the following changes.

Participants

Six graduate students, including one author (HM), participated in the experiment, and all displayed normal color vision. Two of them had participated in Experiment 1.

Procedure

The task involved detection of a switch in a prespecified dimension. In the color condition, participants judged “yes” when a color switch occurred. That is, participants were required to respond “yes” to color-switch and both-switch events and “no” to shape-switch events. In the shape condition, participants were required to respond “yes” to shape-switch and both-switch and “no” to color-switch events. Although the task has two response alternatives as in simple change detection, participants needed to use a complete object file or a selectively maintained partial object file to make a correct judgment. The entire experiment comprised two experimental sessions, each containing 288 trials. Participants performed one session a day. The task-relevant dimension was fixed within each session, and the order of task-relevant dimension was counterbalanced across participants. For each cueing condition, each object motion condition comprised 72 trials, with 24 trials for each event type.

Results and discussion

In the following experiments using a yes–no task, an estimated capacity measure based on Cowan’s $K$ (Cowan, 2001) with an extension for multiple changing objects (Fencsik, Seymour, Mueller, Kieras, & Meyer, 2002) was used as a dependent measure (Appendix A). We chose the Cowan’s $K$ because (1) the formula does not suffer from response bias; and (2) it can be defined when there are two target objects. Figures 5a and 5b show mean estimated capacities as a function of object motion and cueing in the color and shape conditions, respectively. ANOVA with a $2 \times 2 \times 2$ design revealed only a significant main effect of object motion ($F(1,5) = 36.90, p < .001$), with no significant main effect of cueing ($F(1,5) = 0.72, p > .1$). Clearly, postcue did not help detect switches in task-relevant features.

A relevant-feature switch detection task failed to show any effect of postcue, suggesting that the results in Experiment 1 were not simply due to the complexity of response mapping. Selective maintenance of a partial object file representing the task-relevant feature pair leads to a correct judgment. However, the lack of postcue benefit suggests that participants cannot perform such selective maintenance. Note that the lack of postcue benefit in this experiment is somewhat inconsistent with the significant single probe benefit with the color–shape binding condition found by Wheeler and Treisman (2002), because in this condition, benefit was observed even though an irrelevant feature (location) was involved. This issue will be further discussed in the General discussion section.

Overall, the results of Experiments 1 and 2 suggest that the performance deficit for object files in MOPT task performance is due to the maintenance of object files, not their retrieval. To secure this conclusion, however, some alternative interpretations need to be examined. First, the flashing cue used in Experiments 1 and 2 might have been ineffective, unlike the single probe manipulation of Wheeler and Treisman (2002). Second, some factors specific to MOPT task may have eliminated the postcue effect. Third, and most importantly, the relevant-feature switch detection task in Experiment 2, although simpler than the full type-identification task, may still be too complex to show any postcue benefit. To test the

![Figure 5](http://jov.arvojournals.org/pdfaccess.ashx?url=/data/journals/jov/933533/ on 04/07/2017)
plausibility of these alternatives, Experiment 3 used a simple change-detection task with postcue. If the procedure employed in Experiment 2 was responsible for the lack of postcue benefit, the simple change-detection task with the same procedure will again show no postcue benefit. If the lack of postcue benefit is due to the complexity of the relevant-feature switch detection task, the simple change-detection task will show postcue benefit, and at the same time, task performance in general will be improved compared with the relevant-feature switch detection task. Finally, the hypothesis that the retrieval cue benefit found by Wheeler and Treisman is restricted to nonselective processing of partial object files predicts that the simple change-detection task will show postcue benefit, and no significant difference in task difficulty will be evident between Experiments 2 and 3. This is because unlike the relevant-feature switch detection task, the simple change-detection task does not require discrimination of change types and thus can be performed with partial object files.

**Experiment 3**

**Method**

The method was the same as in Experiment 1, except for the following changes.

**Participants**

Six graduate students, including one author (HM), participated in the experiment, and all displayed normal color vision. Four of them had participated in Experiment 2.

**Materials**

No-change sequences were added. Four event types were used: color change, shape change, both change, and no-change. The task was simple change-detection, so no-change was mapped to the “no” response, and all other types were mapped to the “yes” response. The precue condition was eliminated, so all cues were postcue. The cue was 100% valid when a switch occurred. In the no-change sequence, the cue was presented to a randomly selected object, and cue timing was matched to sequences with a switch. To reduce sequence length, each sequence ended two occlusion periods after the switch event. Length of the no-change sequence was matched to those with switch events.

**Procedure**

The task was a simple change detection. The participant judged “yes” when any type of switch was detected. The entire experiment comprised two experimental sessions, each containing 144 trials. One session contained cues and the other did not, and the order of cue and no-cue sessions was counterbalanced across participants.

**Results and discussion**

Figures 6 show estimated Cowan’s $K$ as a function of object motion and cueing. Postcue displayed a facilitatory effect in the stationary condition, but not in the moving condition. ANOVA with a 2 (cueing) × 2 (object motion) design revealed a significant main effect of cueing ($F(1,5) = 6.618, p < .05$), main effect of motion ($F(1,5) = 98.016, p < .001$), and interaction ($F(1,5) = 8.328, p < .05$). Planned comparisons showed that postcue condition displayed a significantly larger estimated capacity in the stationary condition ($F(1,5) = 10.390, p < .05$), whereas no significant effect was seen in the moving condition ($F(1,5) = 2.501, p > .1$). These results suggest that the lack of postcue effects in Experiment 1 can be attributed to limits in the maintenance capacity of complete object file representations.

A simple change-detection task revealed significant facilitation in the stationary condition. A retrieval cue facilitates judgment of whether any kind of change is present but does not help identify the type of switch. The postcue paradigm can thus reveal a facilitation effect similar to that found with the single-probe paradigm of Wheeler and Treisman (2002), suggesting that postcues used in this study can effectively function as a retrieval cue. Moreover, given that the results of Experiment 3 are consistent with previous findings using a single probe paradigm of a change-detection task, the MOPT and change-detection tasks measure some common aspects of VSTM. Furthermore, the task difficulty indexed by Cowan’s $K$ is quite comparable between Experiments 2 and 3, suggesting that the lack of postcue benefit in Experiment 2...
does not reflect task complexity. The pattern of results thus suggests that the cueing benefit observed here and by Wheeler and Treisman may be limited to memory for partial object files when there is no need to know the source of the switch event.

Another interesting result was the lack of postcue effects in the moving condition, suggesting that the postcue is ineffective for moving objects. This may reflect that memory retrieval and matching operation are location based, not object based. The lack of object-based postcue effect provides another piece of evidence against the involvement of complete object files in a simple switch-detection task, as spatiotemporal updating of representation with object motion is a hallmark of object file representation.

Experiments 2 and 3 suggest that retrieval cues are advantageous only for partial object files, not for complete object files, but different results between the two experiments may reflect other differences in experimental methodology. First, task difficulty may be the primary cause of the difference. Estimated capacities in the two experiments were not substantially different but were in fact lower in Experiment 2, and sampling error may have reduced the true difference in task difficulty. Second, the factor of cueing was manipulated in blocks in Experiment 3 but was randomly mixed in Experiments 1 and 2, which may have contributed to the differences in cueing effects. Third, base rates of yes trials differed between Experiment 2 (0.667) and Experiment 3 (0.5), and might thus have subtly influenced participants’ performance. Finally, the flashing cue might have disrupted performance in the relevant-feature detection task more than in the simple detection task. The flashing cue involves a brief color change, which may have impaired color memory processing. If so, the use of the flashing cue might have disrupted performance for the color condition in Experiment 2, in which participants were asked to selectively process colors. In addition, this might have affected performance in the shape condition in Experiment 2, particularly when the shape condition was conducted after the color condition.

Based on these concerns, Experiment 4 directly compared relevant-feature switch detection and simple change-detection tasks using a within-subject design, keeping methodological parameters identical. If the results in Experiments 2 and 3 are reliable and not due to methodological differences, the simple change-detection task will show postcue benefit, whereas the relevant-feature switch-detection task will not show any benefit.

### Experiment 4

#### Method

**Participants**

Six graduate and undergraduate students at Kyoto University participated in the experiment, and all displayed normal color vision. None of the participants had participated in any of the previous experiments.

**Design**

Two independent variables were task (simple change-detection and relevant-feature switch detection) and cueing (postcue and no cue), both of which were within-subject factors. The relevant-feature switch detection task was the same as that in Experiment 2, but only a shape-detection task was used. The white flashing cue in Experiment 2 might have disrupted color memory processing, representing a possible cause of the lack of postcue benefit. The simple change-detection task was the same as that in Experiment 3, except that cueing conditions were randomly mixed within an experimental session. The two tasks were conducted in two separate experimental sessions, conducted on separate days. The order of sessions was counterbalanced across participants.

**Materials**

Several changes in stimulus parameters were seen from Experiments 1–3. First, the moving condition was eliminated. Second, initial angular locations of the four objects were randomized while keeping the four objects 90° apart from each other on an imaginary circle, unlike the fixed initial locations (top, right, bottom, and left) in Experiments 1–3. This manipulation was necessary because without the moving condition, locations of the four objects become completely fixed throughout the experiment, and some participants in a preliminary experiment reported the use of task-specific strategies such as imagining the four objects as four parts of a human face. Visible and invisible periods were both 480 ms, and a flashing cue was presented at 240 ms for 40 ms, slightly shortened from previous experiments. A single trial contained six occlusion periods, reduced from seven in previous experiments.

**Procedure**

The entire experiment comprised two experimental sessions, each containing 144 trials. Participants performed one session a day. Within each session, cueing conditions were randomly mixed from trial to trial. In the relevant-feature switch detection task, each cueing condition comprised 72 trials, with 24 trials for each of the three event types (both-switch, shape-switch, and color-switch). The ratio of yes and no trials was thus 2 to 1. To match the base rate of the yes trials between tasks, in the simple detection task, 72 trials for each cueing condition were divided into 24 no-switch, and 48 switch trials composed of 16 trials for each of the three event types. To avoid verbal encoding of color and shape, articulatory suppression was used by getting subjects to say “coca-cola, coca-cola…”
Results and discussion

Figure 7 shows estimated Cowan’s K as a function of task and cuing conditions in Experiments 1–3. Consistent with Experiments 2 and 3, postcue displayed a facilitatory effect in the simple change-detection condition, but not in the relevant-feature switch-detection condition. ANOVA with a 2 (task) × 2 (cueing) design revealed a significant main effect of cueing ($F(1,5) = 11.432, p < .025$) and an interaction close to statistical significance ($F(1,5) = 6.387, p = .052$). The main effect of task was not significant ($F < 1$). Planned comparisons showed that postcue condition displayed a significantly higher capacity estimate in the simple change-detection condition ($F(1,5) = 51.618, p < .001$), whereas no significant effect was seen in the relevant-feature switch-detection condition ($F(1,5) = 0.002, p > .1$). The results replicated those in Experiments 2 and 3, suggesting that the postcue benefit is observed only when the task allows correct response without discriminating the source of switch event. Also, the two tasks displayed no difference in their difficulty, arguing against the notion that task difficulty was the primary determinant of the postcue benefit in Experiments 1–3.

Experiment 5

One alternative account of the results of Experiments 1–4 is memory overwriting. Memory for triple conjunctions may be maintained in memory but quickly overwritten when features reappear in a new combination. Distinction of maintenance limit and overwriting is theoretically important, because this has to do with the general issue of decay and interference in memory. The postcue provided to a postchange object may hinder the retrieval of prechange representation due to overwriting. To reduce the effect of overwriting, we introduced two new manipulations: elimination of spatiotemporal continuity of pre- and postchange objects and modification of cue timing. In the new occluder-cue condition, the test object was presented without any spatiotemporal overlap with the cued object, and a cue was provided during occlusion by flashing an occluder. The occluder cue was provided before the appearance of postchange object, thus memory retrieval should not suffer from overwriting. Although an occluder may overwrite the representation of the occluded object due to spatial overlap, object files are known to survive occlusion events (Scholl & Pylyshyn, 1999), and a previous experiment with MOPT showed no overwriting effects by occluder (Saiki & Miyatsuji, 2007). Thus, overwriting by an occluder, if any, would be negligible compared with overwriting by feature recombination. The relevant-feature detection task in Experiment 4 was performed both with object-cue (as in Experiment 4) and with occluder-cue. If the lack of the postcue benefit in Experiments 2 and 4 was due to memory overwriting, the occluder-cue condition should display significant improvement compared with the object-cue condition.

Method

Participants

Ten graduate and undergraduate students at Kyoto University participated in the experiment, and all displayed normal color vision. Three of them had participated in Experiment 4.

Design

The independent variable was cueing, comprising object-cue and occluder-cue conditions.

Materials

The object-cue condition was identical to the relevant-feature switch with postcue condition in Experiment 4. In the occluder-cue condition, instead of providing a postcue to the changed object, one blade of occluder was flashed to provide a cue to the object to be tested. Furthermore, to avoid overwriting, stimulus sequence was terminated at the end of the occlusion period with cue, and the test object was presented 6.2° below the center of the occluder, in a region without any overlap with the four objects and occluder. Flashing occluder cue was presented at 240 ms of occlusion for 40 ms, then the occluder stopped rotating and remained on the display for 240 ms. The test object was presented at offset of the occluder display and remained on the display until a response by the participant. This 240-ms temporal offset of cue and test in the occluder-cue condition was set to be comparable to the 240-ms interval between the start of a visible period and postcue in the object-cue condition. According to the overwriting hypothesis, intact memory representa-
tion begins being overwritten from the presentation of changed objects, so the 240-ms period between presentation of changed object and postcue suffers from substantial overwriting. In contrast, in the occluder-cue condition, the 240-ms period between cue and test does not suffer from overwriting. The test object was the changed object at the cued location in the object-cue condition and thus differed in either color, shape, or color-and-shape. A schematic illustration of stimulus sequence is shown in Figure 8.

**Procedure**

The entire experiment comprised two experimental sessions, each containing 72 trials. Within each session, trials with one task condition were performed. At the beginning, instructions for the task were given, followed by eight practice trials. Each cueing condition comprised 72 trials, with 24 trials for each of the three event types (both-switch, shape-switch, and color-switch). The task was identical to the shape switch detection task in Experiment 4, so both-switch and shape-switch trials required a yes response, and color-switch trials required a no response. Only the postcue condition was used, so a cue was provided on every trial. In the occluder-cue condition, participants were instructed to compare the object behind the cued blade and single test object. A short break was given before the second session. The order of experimental sessions was counterbalanced across participants.

**Results and discussion**

Figures 9 shows estimated Cowan’s $K$ as a function of cuing conditions. No significant difference existed between object-cue and occluder-cue conditions ($F(1,9) < 1$). These results do not support the overwriting account, which predicts that the occluder-cue condition improves task performance. The target object was cued during occlusion, and the target and test objects had no spatiotemporal overlap in the occluder-cue condition, so the occluder-cue condition is unlikely to be the target of memory overwriting. Although the null result cannot completely eliminate the possibility of overwriting, we are confident that overwriting does not play a major role in the performances observed in Experiments 1–4. Overall performance was lower in Experiment 5 than in previous experiments, probably reflecting the fact that the number of experimental trials was smaller in Experiment 5 than in previous experiments.

**General discussion**

The present study investigated the nature of limitation in feature-bound memory in VSTM, by examining whether
retrieval cue benefits for memory for feature pairs can be generalized to memory for more complex binding, using a cueing version of MOPT. This issue is theoretically important because object files are often postulated for memory representations of objects in VSTM, but no convincing support for this claim has been offered. The critical piece of evidence should be the involvement of object files in the retrieval cue benefit in VSTM. We addressed this issue by introducing new tasks capable of evaluating the involvement of triple conjunctions.

Overall, a series of four experiments failed to obtain evidence that previous retrieval cue benefit for feature pairs (called partial object files) can also occur for complete object files. A significant effect was seen for precue, but not for postcue, suggesting that helping memory retrieval does not facilitate task performance. This was not due to any peculiarity of the MOPT paradigm itself, or the type of cue, as Experiments 3 and 4 with a simple change-detection version of MOPT showed results consistent with previous findings. Furthermore, the lack of postcue effect cannot be attributed to the complexity of the type-identification paradigm. Experiments 2 and 4 with a relevant-feature switch detection task also failed to obtain any postcue effect, suggesting that no postcue effect was observed when two different types of switch events had to be discriminated. Comparable task difficulty between the simple change-detection and the relevant-feature switch-detection tasks in Experiments 2–4 provides further evidence against the possibility that a lack of postcue benefit reflects task complexity or task difficulty. Finally, the results in Experiment 5 suggest that these findings are unlikely to reflect memory overwriting. Taken together, the interaction between postcue benefit and task (significant benefit with the simple change-detection and no benefit with tasks requiring triple conjunctions) suggests that retrieval cue benefit occurs only for partial object files, and that limits in complete object files primarily reflect memory maintenance. Maintenance capacity for complete object files is close to the estimated capacity, that is, one or two objects, whereas that for partial object files may be larger. Below, we discuss the validity and generality of these findings and the nature of memory representations for feature binding in VSTM.

**Validity and generality of the current findings**

One may argue that the findings of this study using the MOPT paradigm are specific to the experimental paradigm and not generalizable to more typical tasks such as change detection. First, the MOPT task is arguably more complex than a simple change-detection task. One should note, however, that the observers’ performance was actually comparable to performance typically observed with change-detection tasks. Accuracy levels in Experiments 2 and 3 were similar to those seen in the binding conditions for Wheeler and Treisman (2002) and other studies. Second, use of the occluder may make the task more difficult by masking objects by the occluder. As Movies 1 and 2 shows, the time course of the MOPT sequence does not impose any strong perceptual masking. Furthermore, Saiki and Miyatsuji (2007) have already examined interference of the occluder by comparing figures with and without occluder and showed no interference effect. Use of the occluder in MOPT is thus unlikely to be responsible for the observed results regarding the effect of postcue.

What is the underlying mechanism of postcue benefit observed with the simple change-detection tasks in Experiments 3 and 4? As stated in the Introduction, the current experiments cannot distinguish memory retrieval per se from comparison of memorial and perceptual representations. The postcue thus either improves retrieval of cued item itself or improves the comparison process by selective encoding of the cued item. One extreme alternative is that postcue does not affect memory contents or access and retrieval of these contents at all, but simply improves perceptual processing of the cued object. Findings of the current study and previous work have been unable to distinguish between these alternatives, and further experiments are necessary.

Whichever the mechanism of postcue benefit, however, the presence and absence of postcue benefit provides important theoretical insights into the nature of memory for feature binding. Even if the postcue exclusively affects perceptual processing of the cued object, the main theoretical claim of this study still holds. This is because facilitation of perceptual processing of the cued object affects matching operation of memory and perceptual representations, not maintenance capacity per se. Thus, the significant postcue benefit with the simple change-detection task implies that maintenance capacity in the no-cue condition was underestimated by matching cost. The lack of postcue benefit with tasks requiring triple conjunctions implies that estimated capacity in the no-cue condition reflects maintenance capacity. The main theoretical argument of this study, that complete and partial object files are not functionally equivalent, is thus independent of the underlying mechanism of postcue benefit.

As stated in Experiment 2, results in the relevant-feature switch-detection tasks were somewhat inconsistent with the shape-and-color binding experiment of Wheeler and Treisman (2002). In their experiment, locations were irrelevant and location changes should be ignored, but a significant benefit of single probe was observed. Filtering out color or shape from triple conjunction of color, shape, and location thus appears difficult, whereas location is easier to filter out, suggesting different functional properties between location-based partial object files and shape–color partial object files. However, the relevant-feature switch detection task in this study and change-detection task of Wheeler and Treisman display important methodological differences. For example, the color–shape
conjunction experiment used random variation of irrelevant feature, whereas relevant-feature switch-detection task used a systematic change of irrelevant features. More systematic comparisons are thus needed to understand the mechanisms underlying differential effects of irrelevant features.

**Nature of memory representations for feature binding**

The present results argue against the view that memory of feature binding is a system composed of complete object file representations. Complete object file representations should lead to postcue benefits in all different tasks used in this work, a possibility that was unsupported by the data. Unlike a previous claim by Luck and Vogel (1997) that the content of object memory, object files, is complete, regardless of the number of features, the present study suggests that the content of object files are partial by default. The present study suggests that functional properties of object files differ depending on complexity, which is related to a recent argument regarding whether complexity of objects affects the capacity of VSTM (Alvarez & Cavanagh, 2004; Awh, Barton, & Vogel, 2007).

Alvarez and Cavanagh (2004) reported that capacity estimate using a simple change-detection task is a linear function of the complexity of the object measured by the slope in a visual search task, suggesting that the complexity of objects affects the capacity of VSTM. Recently, however, Awh et al. (2007) showed some evidence that these results could be explained by difficulty of matching between memory and percept, suggesting that the capacity of VSTM is fixed regardless of object complexity, but resolution of object representations becomes degraded with increasing complexity. As far as the simple change-detection task is concerned, the results of the present study appear consistent with the argument by Awh et al. as the significant postcue benefit with simple change detection suggests that performance impairment primarily reflects memory retrieval or matching of memory and percept, and not capacity per se. The idea that increasing complexity of the object degrades resolution of object representation can be interpreted as meaning that complex objects are represented as an unselected collection of partial object files, which are presumably less veridical descriptions of the object compared with a complete object file. In contrast, the results with tasks requiring triple conjunctions seem consistent with the argument of Alvarez and Cavanagh, suggesting that impairment primarily reflects maintenance capacity. When the task requires use of complete object files, the capacity of object file representation is substantially reduced. Taken together, the idea of fixed capacity with varied resolution may hold only in the context of simple change detection, and in general, the complexity of objects may reduce the maintenance capacity of memory representation.

At least two functional architectures of memory systems can account for the above findings. One involves a single memory system for object files, and the other involves separate memory systems for partial and complete object files. The single memory system view needs to assume that, depending on the task, the system forms either partial or complete object files. For simple change detection where an unselected collection of partial object files is sufficient, the system forms partial object files with relatively large maintenance capacity. For tasks requiring complete object files, maintenance capacity will be reduced. Thus, not stimulus complexity alone, but complexity in conjunction with task requirement affects maintenance capacity of the memory system.

An alternative architecture with separate systems for complete and partial object files assumes that complete and partial object files are stored more or less independently. A partial object file system may be responsible for performance in simple change-detection tasks, offer relatively large fixed capacity with resolution that decreases with object complexity, and be vulnerable to memory retrieval bottlenecks. Consistent with Wheeler and Treisman (2002), the capacity of a partial object file is underestimated by simple change-detection tasks. A complete object file system, evaluated by type-identification and relevant-feature switch-detection tasks, appears to have a highly limited capacity, and this limit reflects memory maintenance, not memory retrieval. Provided that partial object files are insufficient for type-identification and a relevant-feature switch-detection task, performance in these tasks primarily reflects properties of the complete object file system. In contrast, performance in a simple change-detection task may reflect both complete and partial object files. However, given highly limited capacity, the complete object file system seems unlikely to contribute to the capacity limit or benefit of retrieval cueing found in the simple change-detection task.

One recent study suggests multiple memory systems for object VSTM (Xu & Chun, 2006). Using functional magnetic resonance imaging (fMRI) experiments, Xu and Chun (2006) found two regions sensitive to capacity of VSTM: the inferior intraparietal sulcus (IPS), which represents a fixed number of objects regardless of complexity; and the superior intraparietal sulcus, which represents objects for which capacity varies with object complexity. The inferior and superior IPS may correspond to partial and complex object file systems, respectively.

If separate systems exist for object VSTM, what is the representational scheme for the partial object file system, which is sufficient for simple change detection? One possibility is a feature map representation. Suppose each object feature (color and shape) is bound to the location, forming a color map and shape map, independently. If these feature maps have a capacity of three to five objects and are vulnerable to retrieval bottleneck, postcue benefit in the simple change-detection task can be readily explained, as detection of change by either
map is sufficient. An alternative representational scheme is the saliency map (Itti & Koch, 2000; Koch & Ullman, 1985). Salience representations mix up feature information to represent salience of a stimulus and thus do not explicitly represent the combination of features. However, discriminating whether two displays are the same or different may be sufficient. Further studies are necessary to examine the plausibility of these schemes as VSTM subsystems.

The highly limited capacity for complete object file system may be specific to explicit measure of VSTM. The lack of postcue benefit in type-identification and relevant-feature switch-detection tasks revealed that postcue does not facilitate retrieval of complete object file representations in explicit tasks. One possibility is that the complete object file system of VSTM may have a capacity comparable to the partial object file system, but multiple complete object files may only be available in implicit ways. Although some studies have used implicit measures such as an object-review paradigm (Kahneman et al., 1992), no clear evidence supports the existence of higher-order conjunction representations. Exploration of feature-bound representations in implicit VSTM is an important future direction to resolve this issue.

Finally, what are the theoretical implications of the finding that probably only one complete object file can be maintained? When we try to maintain multiple objects in VSTM, what we maintain are probably three to five partial object files, which is sufficient to just detect a change, but insufficient to judge how they change. To correctly judge how they change, complete object files are necessary, and their maintenance seems to require substantial additional processing costs, reducing the maintenance capacity. Therefore, difference in capacity estimate between the present and previous studies can be explained by difference in task demand; to just detect a change or to judge how they change.

Although several issues remain unresolved regarding the nature of object file representations in VSTM, the present results showed that manipulation of task requirement in combination with stimuli provides a promising method for elucidating the nature of object file representations. In particular, comparisons of performance across different tasks allow the effects of complete object files to be distinguished from those of partial object files. In future studies, behavioral and functional brain imaging experiments with such task manipulations may be able to reveal brain mechanisms underlying different memory systems for objects.

**Conclusion**

The effects of retrieval cue on visual short-term memory depend on task requirement. Whereas a simple change-detection task shows a facilitatory effect as seen in previous studies, tasks requiring discrimination of different feature combinations failed to show facilitation, even when task difficulty was similar to the change detection. These results suggest that retrieval cue benefit occurs in memory for partial object files, but not for complete object files. Limits in memory for complete object files primarily reflect maintenance capacity, whereas maintenance capacity for partial object files is underestimated by a simple change-detection task due to retrieval bottleneck.

### Appendix A

**Modified Cowan’s K**

The original Cowan’s $K$ index is defined for a change-detection task with a single changing item, which is:

$$K = (h + cr - 1)N,$$

where $N$, $h$, and $cr$ are the number of items, hit rate, and correct rejection rate, respectively. Conceptually, this is:

$$h + cr - 1 = K/N = p_{\text{store}},$$

where $p_{\text{store}}$ is the probability of storing the to-be-switched object.

With one changing target, the postulation of $p_{\text{store}}$ being $K/N$ is valid under the high-threshold theory, but when features of two items switch contains two changing items as in this study, an appropriate modification is necessary. Fencsik et al. (2002) proposed a modification formula for Pashler’s index, where $p_{\text{store}}$ is defined as:

$$p_{\text{store}} = 1 - \prod_{i=0}^{c-1} \frac{N - k - i}{N - i},$$

where $c$ and $k$ are the number of changing objects and capacity estimate, respectively. In the case of this study, where $N$ is 4 and $c$ is 2, combining Equations (A2) and (A3) leads to:

$$k = \frac{7 - \sqrt{49 - 48(h + cr - 1)}}{2}.$$  

Note that Equation (A4) can properly estimate the maximum capacity in the case of $N$ being 4 and $c$ being 2 as 3, meaning that when 2 objects change out of 4, storage of 3 objects guarantees perfect performance, whereas the original Cowan’s $K$ overestimates maximum capacity as 4.
Acknowledgments

This work was supported by a Grant-in-Aid for scientific research from JMEXT (#19500226), PRESTO “Intelligent cooperation and control” from Japanese Science and Technology Agency (JST), the 21st Century COE Program from JMEXT (D-2 to Kyoto University), and the Global COE Program “Revitalizing Education for Dynamic Hearts and Minds,” MEXT, Japan. We wish to thank Alex Holcombe for helpful comments on earlier drafts of this manuscript.

Commercial relationships: none.
Corresponding author: Jun Saiki.
Email: saiki@cv.jinkan.kyoto-u.ac.jp.
Address: Graduate School of Human and Environmental Studies, Kyoto University, Yoshida-Nihonmatsucho, Sakyo-ku, Kyoto, 606-8501, Japan.

References