Ψ Psychology Press
Taylor & Francis Group

Overwriting and rebinding: Why feature-switch detection tasks underestimate the binding capacity of visual working memory

George A. Alvarez and Todd W. Thompson

Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology, Cambridge, MA, USA

In these two experiments, we explored the ability to store bound representations of colour and location information in visual working memory using three different tasks. In the location-cue task, we probed how well colour information could be recalled when observers are given a location cue. In the feature-cue task, we probed how well location information could be recalled when observers are given a colour cue. Finally, in the feature-switch detection task, we tested how well observers could detect a recombination of features (e.g., switching the locations of the red and green items). We hypothesized that these tasks might reveal differences in binding capacity limits between switching and nonswitching tests of visual working memory. We also hoped the tasks could provide an explanation for those differences in terms of the component processes of working memory—do failures occur in the encoding, maintenance, or retrieval stages of the task? Experiment 1 showed that performance in the two cued-recall tasks was equally high, and was significantly better than performance in the feature-switch detection task. Thus, the feature-switch detection task underestimates the number of colour–location bindings that can be remembered, but is a useful task for examining the fragile nature of feature binding in working memory. Experiment 2 explored why feature-switch detection underestimates the binding capacity of visual working memory by examining whether the feature switch errors occur at the level of encoding, maintaining, or retrieving binding information from visual working memory. The results suggest that feature switch errors reflect failures to maintain bound objects in working memory, perhaps due to the automatic rewriting and rebinding of information in the face of new perceptual input.

Keywords: Working memory; Binding; Overwriting; Cued recall; Feature switch.

Perhaps the most basic computation the human visual system carries out is parsing the incoming stream of visual information into meaningful units, such as surfaces, objects, and events. The fact that this seems to be accomplished with little conscious effort belies the challenge the visual system faces of integrating many independent sources of information into these cohesive units. As noted by many previous researchers, the visual system seems to be comprised of specialized, parallel processing streams for different visual feature dimensions (e.g., colour, motion, etc.; for reviews see Cowey, 1985; Zeki, 1993). This gives rise to the perceptual "binding problem": If these features are analysed in separate brain regions, how are they subsequently combined to form meaningful, integrated units? Research on this perceptual binding problem has supported the hypothesis that attention is needed to integrate the features of an object (Treisman, 1996, 1998). Attending to an object binds the features of that object together, whereas withdrawing attention from an object will result in a loss of binding information (Wolfe, Klempen, & Dahlen, 2000) or binding errors such as illusory conjunctions (Treisman & Schmidt, 1982).

In contrast with these previous studies that have investigated the perceptual binding problem, the current study investigates visual working memory for bound objects. However, it is important to consider first whether there exists a binding problem in visual working memory. How can we determine whether errors in a visual working memory experiment reflect true binding errors, or if they simply indicate a failure to retrieve a component feature of a successfully bound object? There would be a stronger case for a specific "binding problem" in visual working memory if it were possible to store the constituent features of an object without remembering how they were bound together. This has in fact been shown in a previous study by Stefurak and Boynton (1986), who showed that it is possible for observers to remember the colours and shapes of objects that were seen, without remembering which colour was bound to which shape. These results suggest that binding operations, and the maintenance of bound object representations, are part of a distinct process or level of representation that can be isolated for further investigation.

One of the central issues addressed in this paper is a methodological question: Do different methods of probing the contents of visual working memory reveal different aspects of binding and feature storage? A variety of methods is employed in the literature for probing visual working memory, including recall, old/new recognition, and change detection. In each of these tasks, a sample display consisting of several multidimensional objects is briefly presented, followed by a retention interval on the order of seconds, which is then followed by a test. A typical recall test would require observers to report the identity of all objects. In cued-recall one feature of the test items is used as a cue to probe recall for other features of that item (e.g.,

what colour was the square, with shape serving as a cue for colour recall). In old/new recognition tasks, a single test item is presented, and the observer must report whether that item was a member of the memory set. Finally, in change detection, the test display is a second array of objects. On half of the trials all of the items appear identical to the sample display, and on the other half of the trials a single item is different in some way (e.g., in colour, shape, etc.). The task is to report whether the two displays are the same or different.

Change detection is perhaps the most commonly employed task for investigating the capacity of visual working memory (Alvarez & Cavanagh, 2004; Luck & Vogel, 1997). In a seminal study on visual working memory, Luck and Vogel showed that the number of objects that can be remembered is about four, whether observers are required to remember one, two, or even four features per object (Luck & Vogel, 1997). These results were consistent with their claim that visual working memory stores integrated object representations. However, other researchers have noted that in the typical change detection task, features do not have to be bound together to detect a change (Wheeler & Treisman, 2002). In the typical change detection task, only a single feature is probed at test, even when multiple features such as colour and orientation are remembered (Luck & Vogel, 1997). If there were completely independent feature stores for colour and orientation, then a change to either feature dimension could be detected even if there was no binding across dimensions.

To address this limitation of the standard change detection task for investigating feature binding in memory, Wheeler and Treisman (2002) introduced a modified change-detection task, which we will refer to as *feature-switch detection*, in which observers must look for a feature switch. For example, if the task were to remember combinations of colour and shape, then a set of objects with different shapes and colours would be briefly presented (e.g., a red circle, blue square, and green triangle), followed by a brief retention interval, and then by a test display. On half of the trials, the test display would be identical to the sample display. On the other half, two objects would switch features (e.g., the red circle and blue square might change to a blue circle and red square). The task is to determine whether or not any of the features have switched. It is impossible to perform this task simply by remembering which features were presented, because all of the same features appear in both the sample and test displays. Thus, it is necessary to remember how the features were combined to detect a feature switch.

Results using this feature-switch detection task and a dynamic variant in which objects move about the display (Saiki, 2002) have shown that feature switch detection is quite poor (Saiki, 2002, 2003a,2003b; Treisman & Zhang, 2006; Wheeler & Treisman, 2002). The difficulty of feature-switch detection could be due to failures at any stage of working memory, including the encoding, maintenance, or retrieval of information (Atkinson & Shiffrin,

1968). Saiki (2002, 2003b) has interpreted the failure to detect feature switches as evidence for encoding failures, and suggested that working memory storage capacity for feature bindings is more limited than for independent features. Other researchers interpret the difficulty of feature switch detection as evidence for maintenance failures, suggesting that when attention is withdrawn from active maintenance, the stored binding information is lost (Wheeler & Treisman, 2002). Finally, the change detection literature suggests that feature-switch detection errors could also occur due to a failure to retrieve information from visual working memory (Mitroff, Simons, & Levin, 2004; Simons, 2000; Simons, Chabris, Schnur, & Levin, 2002).

Because previous research suggests that feature-switch detection tasks are more difficult than other types of task, such as change detection (Wheeler & Treisman, 2002), we hypothesized that feature-switch detection underestimates the binding capacity of visual working memory. In Experiment 1, we investigated the extent to which the feature-switch detection task underestimates working memory capacity by comparing performance on feature-switch detection to performance on cued-recall within the same group of participants. The results showed that feature-switch detection does indeed underestimate the binding capacity of visual working memory. As already noted, the feature-switch performance alone cannot be used to isolate whether feature-switch errors are due to encoding, maintenance, or retrieval failures. As described by Gajewski and Brockmole (2006), the feature-switch detection task alone enables us to determine when a memory failure occurs, but it does not enable us to determine the nature and extent of the failure (i.e., was the entire object forgotten, or was the binding between two features lost?). In contrast, the cued-recall task enables us to assess exactly what type of memory failure occurs by evaluating which features observers report when they make errors. Thus, Experiment 2 explored whether feature-switch detection errors reflect failures of encoding, maintenance, or retrieval by having observers perform the feature-switch detection task and a cued-recall tasks on the same trial. The results showed that the contents of memory are rapidly overwritten by new perceptual input, suggesting that feature switch detection tasks introduce an additional load at the maintenance stage of visual working memory.

EXPERIMENT 1: FEATURE-SWITCH DETECTION VERSUS CUED RECALL

To test whether the feature-switch task underestimates the binding capacity of visual working memory relative to cued-recall tasks, we had observers remember the binding of colour and location, using a dynamic feature-switch detection task modelled after Saiki's (2002) task. Multiple items of different

colours rotated around a fixation point, disappearing and reappearing behind an occluding surface (see Figure 1). This dynamic display places high demands on operations that bind colour to location, as the colour–location combinations are constantly updating, requiring observers to continuously encode
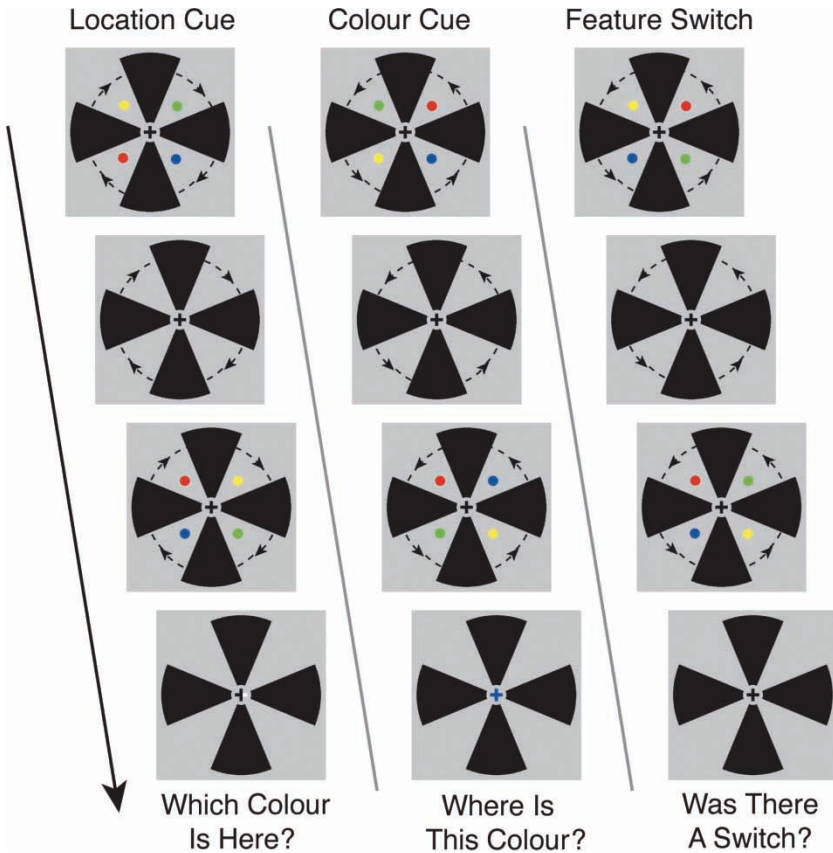

Figure 1. Sample trials for the three different conditions. In each condition, a red, green, blue, and yellow circle rotated either clockwise or counterclockwise (direction chosen randomly for each trial), disappearing and reappearing behind the black occluding surfaces. The task was to keep track of which colour was in which location. In the location-cue condition, all of the items stopped behind the occluders, and one arm of the fixation "+" turned white, prompting the observer to press a key to indicate which colour was in the cued location. In the colour-cue condition, all of the items stopped behind the occluders, and the fixation "+" changed red, green, blue, or yellow, prompting the observer to press an arrow key to indicate where that colour was located. In the feature-switch condition, there was a 50/50 chance that two of the colours would switch places with each other during the trial. In this example, the yellow and blue circles switch places with each other when they reappeared. At the end of the trial, all of the items stopped behind the occluders, and observers indicated whether or not they had detected a switch.

information into visual working memory. Observers monitored these displays and attempted to remember which colour appeared in which location, and simultaneously performed a verbal suppression task to minimize verbal encoding of the stimuli. At the end of each trial, all of the items stopped while hidden behind the occluders. We then probed how well observers were able to remember which colour was in which location using three different tasks (in separate blocks of trials). In the location-cue condition, part of the fixation cross changed white, pointing to one of occluders (see Figure 1a), and the task was to indicate which colour was behind that occluder by pressing the keyboard key corresponding to the colour. In the colour-cue condition, the fixation point turned red, green, blue, or yellow (see Figure 1b), and the task was to indicate where that colour was located by pressing the arrow key corresponding to the correct occluder position (up, down, left, or right). In the feature-switch condition, the task was to report whether any two of the colours had switched locations with each other during the trial. In this condition, 50% of the trials had a feature-switch in which two of the colours switched positions with each other during a disappearance behind the occluders (see Figure 1c). These feature switches only occurred in the feature-switch condition, and occurred only once during a trial. The new feature combination was visible for only a single eighth of a revolution (500 ms). In each condition, observers gave their response to the visual memory task, and then gave their response for the verbal suppression task, typing the letters they had rehearsed throughout the trial. Because these conditions have different levels of chance performance (25% in the cued-recall tasks, and 50% in the feature-switch task), performance was adjusted for chance to enable comparisons across conditions (for more details, see the Data Analysis section).

As described earlier, previous experiments have shown poor performance in this type of feature-switch detection task (e.g., Saiki, 2003a; Wheeler & Treisman, 2002). Because the cued-recall tasks and the feature-switch detection task are identical during the encoding stage of the experiment (see Figure 1), any differences between these tasks must occur during the maintenance or retrieval stages. Thus, to the extent that the feature-switch detection task introduces maintenance or retrieval errors that do not occur in the cued-recall tasks, performance should be worse on the feature-switch detection task (after correcting for differences in chance performance).

Method

Participants

Author GA and 11 naïve observers participated in the experiment. All observers gave informed consent, had 20/20 corrected vision or better, and had normal colour vision.

Apparatus

Experiments were run using the Psychophysics Toolbox extensions (Brainard, 1997; Pelli, 1997), and were conducted on a $35° \times 28°$ CRT monitor running at 60 Hz, viewed from approximately 57 cm. All motion appeared smooth and continuous on this display.

Stimuli

The stimuli were based on a stimulus developed by Saiki (2002) to test binding in visual working memory. As shown in Figure 1, the display consisted of four circles (diameter $= 1°$), each of a different colour (red, green, blue, or yellow), equally spaced on an imaginary ring at an eccentricity of $4°$. Four occluding "wedges" were drawn by dividing a black circle (diameter $= 16°$) into eight sectors, erasing a small circular region near the middle (diameter $= 2°$), and erasing every other sector, leaving the north, east, south, and west sectors to occlude the display. A fixation " $+$ " ($1.5° \times 1.5°$) was drawn in black at the centre of the display. On each trial, the small coloured circles rotated at 0.25 revolutions per second, disappearing behind the black wedges, and then reappearing in the gray zones in between. The display rotated for a random total of one to two cycles (in one-eighth of a cycle increments), and always stopped with the circles hidden behind the black occluders. When there was a feature switch, the new feature combination was visible for a single eighth of a revolution (500 ms) and then the circles. The RGB lookup table values of the colours used in these displays were as follows: Red (255,0,0), green (0,245,0), blue (0,0,255), yellow (255,255,0), black (0,0,0), and grey (185,185,185).

Procedure

At the beginning of each trial, two consonants were presented on the screen and the observer began rehearsing the letters, then pressed a key to begin the visual memory trial. This verbal suppression task has been used in previous visual working memory studies to minimize verbal encoding of visual stimuli (e.g., Gajewski & Brockmole, 2006; Vogel, Woodman, & Luck, 2001). During the trial, four circles of different colours (red, green, blue, and yellow) rotated around the fixation point, disappearing and reappearing from behind occluders, eventually stopping while behind the occluders. The observers' task was to keep track of where each colour was located throughout the trial. Memory for colour–location binding was tested in the three different conditions described earlier and shown in Figure 1, in separate blocks of 12 practice trials and 36 test trials, with the order of conditions counterbalanced across observers (switch, location, colour; colour, switch, location; location, colour, switch).

Data analysis

Due to differences in chance performance for the 4AFC tasks (the colour-cue and the location-cue conditions), and the 2AFC task (the feature-switch condition), we computed three different measures which correct for differences in chance performance: (1) Normalized accuracy, (2) sensitivity in d', and (3) a capacity estimate in terms of the number of objects stored, K.

Normalized accuracy.   This measure takes the raw percentage correct score and rescales it so that chance performance gives a score of zero, and perfect performance gives a score of 100.

$$NA = (PC - chance)/(100 - chance) \qquad (1)$$

Where NA is normalized accuracy, PC is the raw percentage correct score computed directly from the observers' performance, and chance is the chance level of accuracy. An example illustrates how this normalization procedure equates performance for conditions with different chance levels. If Condition A has a chance level of 25% accuracy, then a score of 62.5% would represent a performance level halfway between chance and perfect performance, and would give a normalized score of 50%. If Condition B has a chance level of 50%, then a score of 75% would represent a performance level halfway between chance and perfect performance, and would also give a normalized score of 50%. Thus, according to this procedure, 62.5% correct when chance is 25% is equivalent to 75% correct when chance is 50%.

Sensitivity d'.   We used the standard calculation for d' in the feature-switch condition ($z(HIT) - z(FA)$), where HIT corresponds to the hit rate (probability reporting a feature switch when there was in fact a feature switch), and FA corresponds to the false alarm rate (probability of reporting a feature switch when there was not a feature switch). In the location-cue and colour-cue conditions, d' was calculated based on percentage correct (PC) and the number of response alternatives (M), using the algorithm described by Smith (1982).

Capacity K.   We computed capacity in terms of the number of objects stored (K), based on percentage correct on each task (for similar calculations see Cowan, 2001; Pashler, 1988). In the 4AFC location-cue and colour-cue conditions, percentage correct was assumed to be related to K by the following equation:

$$PC = K/N*1 + (N-K)/N*1/(N-K) \qquad (2)$$

Where PC represents percent correct, K is the number of objects remembered, and N is the total number of items to be remembered. This equation assumes that whenever a remembered item is tested (K/N proportion of the

trials), performance will be perfect, but that whenever an item that was not remembered is tested, performance will be at chance guessing. For example, if an observer can remember two out of four items, then percentage correct should be $2/4*1+(4-2)/4 * 1/(4-2) = .75$. This equation assumes that observers perfectly use the information they have remembered, and can be simplified and solved for K as follows:

$$K_{4AFC} = PC * M - 1 \qquad (3)$$

For the 2AFC feature-switch condition, the intuition for deriving a K score is that subjects will detect a switch if a remembered item is involved in the switch, so given the probability of detecting a switch, we can solve for the number of objects remembered. Thus, if switch detection is limited only by the number of objects remembered, the HIT rate (reporting a switch when there was one) is related to the number of objects remembered (K) and the guessing rate (or FA, the probability of reporting a switch when there was not a switch) by the following set of equations:

$$T = (N*N-N)/2 \qquad (4)$$
$$I = K*N - K*(K+1)/2 \qquad (5)$$
$$HIT = I/T* 1 + FA \qquad (6)$$

Where T is the total number of switches that can occur between pairs of items, and N is the number of items. For four items, there are six possible switches between pairs of items. I is the number of switches that would involve a remembered item if K items were remembered. For example, if only one item out of four items were remembered, $1*4-1*(1+1)/2 = 3$, indicating that three out of the six possible changes would involve that single remembered item. Thus, 3/6 switches would be detected, and the HIT rate would equal to 3/6 plus the false alarm rate. Given the HIT rate, false alarm rate, and number of items to be remembered, these equations can be simplified to the following:

$$0 = -(K^2) + (2*N + 1)*K - 2*T*(HIT-FA) \qquad (7)$$

In this form (zero $= aK^2 + bK + c$), we can solve for K using the quadratic equation, as follows:

$a = -1$
$b = 2*N + 1$
$c = -2*T*(HIT-FA)$
$$K_{4AFC} = [-b +/- sqrt(b^2 - 4*a*c)]/(2*a) \qquad (8)$$

There are two solutions to this equation, but one of them will always be outside the range of possible values for K (less than zero, or greater than N), and so the solution that is inside the range of possible K values was taken as the capacity estimate.

Results

Overall, performance on the verbal suppression task was greater than 96% ($SEM = 2.2\%$), and did not differ significantly as a function of the test condition (all $ts < 1.04$, all $ps > .31$). The remaining analyses were conducted on visual memory performance for trials in which both letters of the verbal suppression task were correctly reported.

Raw accuracy.   Overall accuracy in the 4AFC colour-cue trials was 83% ($SEM = 4.6\%$), indicating that observers could report the location of a particular colour with fairly high accuracy. Similarly, accuracy in the 4AFC location-cue trials was 84% ($SEM = 4.4\%$), indicating that observers could also report the colour at a cued location with fairly high accuracy. Combined, performance in these two conditions suggests that observers know with high accuracy which colours are bound to which locations. In the 2AFC feature-switch task, observers performed at 79% ($SEM = 3.8\%$). However, due to the difference in chance accuracy (25% for the 4AFC tasks, and 50% in the 2AFC task), these raw accuracy scores cannot be directly compared. To account for this, we computed three performance measures that correct for these differences in baseline accuracy: Normalized accuracy, d', and the number of objects stored (see Method). Comparisons with all three measures indicate that memory performance is worse in the *feature-switch condition than in either the location-cue or colour-cue* conditions.

Normalized accuracy.   Normalized accuracy was computed for each condition, as described in the Method section. As shown in Figure 2a, performance was not significantly different in the colour-cue and location-cue conditions, $t(11) < 1$, $p = .463$, $r^2 = .050$, but performance in both of these conditions was better than in the feature-switch condition: Colour-cue vs. feature-switch), $t(11) = 4.19$, $p < .01$, $r^2 = .615$; location-cue vs. feature-switch, $t(11) = 5.14$, $p < .001$, $r^2 = .706$.

d'.   Sensitivity in d' units was computed for each condition, as described in the Method section. As shown in Figure 2b, performance was not significantly different in the colour-cue and location-cue condition ($t < 1$), but performance in both of these conditions was better than in the feature-switch condition: Colour-cue vs. feature-switch, $t(11) = 2.33$, $p < .05$, $r^2 = .331$; location-cue vs. feature-switch, $t(11) = 2.86$, $p < .05$, $r^2 = .427$.

Number of objects stored (K).   An estimate of the number of objects stored (K) was computed for each condition, as described in the Method section. As shown in Figure 2c, performance was not significantly different in the colour-cue and location-cue condition ($t < 1$), but performance in
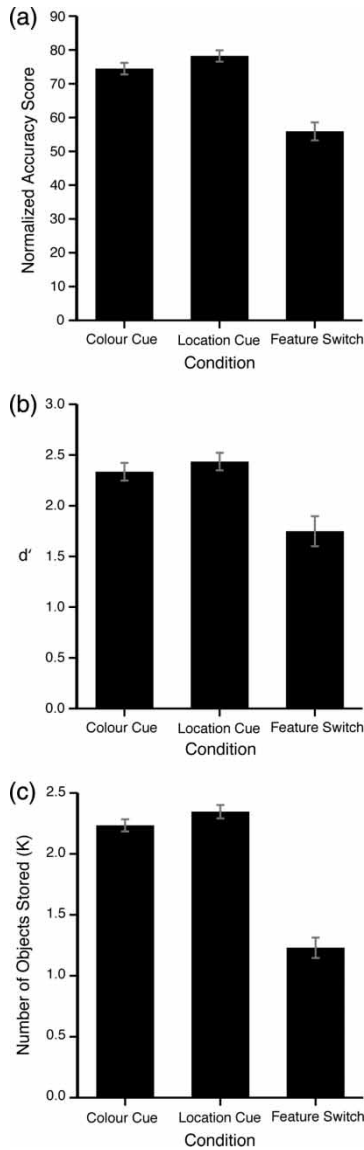
Figure 2.   Results of Experiment 1 expressed in three different measures that correct for differences in chance accuracy: (a) Normalized accuracy scores, (b) sensitivity in terms of d', and (c) number of objects stored, K. Each measure shows that performance was not significantly different in the location-cue and colour-cue conditions, but performance in both of these recall conditions was significantly better than in the feature-switch condition. Error bars represent one within-subject standard error of the mean, calculated with Cousineau's (2005) modification of Loftus and Masson's (1994) method.

both of these conditions was better than in the feature-switch condition: Colour-cue vs. feature-switch, $t(11) = 7.47$, $p < .001$, $r^2 = .835$; location-cue vs. feature-switch, $t(11) = 8.45$, $p < .001$, $r^2 = .866$.

Discussion

This experiment tested how well observers could maintain bound representations of colour and location in visual working memory. In the location-cue and feature-cue conditions, which were both cued-recall tasks, observers could report which colour appeared in which location with high accuracy. However, in the feature-switch condition, performance was significantly worse than in either of the cued-recall conditions. In fact, the switch detection task underestimates the number of colour–location pairs that are initially bound and maintained in visual working memory by approximately 90% relative to the cued-recall conditions.

Previous work has interpreted feature-switch errors as evidence that working memory has a relatively limited capacity to store colour–location combinations (Saiki, 2002; Saiki & Miyatsuji, 2007). Although we replicate the difficulty of feature-switch detection, the high level of cued-recall performance indicates that the failure to detect feature switches cannot be attributed entirely to a failure to encode bound features into visual working memory. It appears that the feature switch detection task places additional demands on the maintenance or retrieval of bound objects from visual working memory.

One possible reason observers might miss a feature switch is that the initially bound object representations might not be maintained after a feature-switch occurs. The dynamic displays we employed required observers to constantly update their binding memory, since the colours are constantly changing location. It is possible that when the items are attended, their features are automatically rebound in memory to match the new perceptual input (the *maintenance-failure hypothesis*). Such automatic rebinding could occur if visible stimuli are given priority over the current contents of memory, or if visible stimuli simply have a more robust representation than those maintained in working memory. In fact, previous research suggests that bound representations are fragile and can be interfered with by subsequent encoding of information (Allen, Baddeley, & Hitch, 2006), or by the diffusion of attention during retrieval (Wheeler & Treisman, 2002). We suggest that, due to the fragile nature of these bound representations, the contents of working memory might be particularly susceptible to an automatic rewriting operation. On this view, objects are initially encoded into memory as bound, integrated units, but the contents of visual working

memory are automatically rewritten when they are perceived in a new combination.

Another possibility is that the initial binding representation was maintained, but was not retrieved due to time constraints or decision complexity (the *retrieval-failure hypothesis*). Evidence for this type of memory access failure has been observed in change-blindness studies (Mitroff et al., 2004; Simons et al., 2002). In these studies, observers often failed to notice changes to the appearance of objects in a display, and yet were able to answer forced choice questions about what the objects looked like before the changes occurred, as if they had a memory representation which simply had not been retrieved. There are multiple reasons similar memory access failures might occur in the feature-switch detection task. Because all the circles were constantly changing location, observers were required to check each item for a switch each time it reappeared from behind an occluder. Even with perfectly bound representations of each item, this could be difficult if there was not enough time to check each item, or if the decision was simply so complex that observers did not exhaustively check each item. The cued-recall tasks do not share these time constraints, or decision complexity, because only one item must be accessed, via either its location or its colour.

Experiment 2 contrasts the maintenance-failure hypothesis with the retrieval-failure hypothesis directly by requiring observers to perform the feature-switch detection task and a cued-recall task within the same trial.

EXPERIMENT 2: EVIDENCE FOR REWRITING AND REBINDING OF WORKING MEMORY CONTENTS

This experiment explored whether feature switches are missed due to a failure to retrieve information from memory, or due to a failure to maintain the original bindings, perhaps due to an automatic overwriting and rebinding of the contents of memory. To address this question, observers performed both the feature-switch detection task and the cued-recall tasks within the same trial. This enabled us to determine the colour–location binding that observers report when they miss a feature switch. The retrieval-failure hypothesis holds that feature switches are missed because there is not enough time, or it is too difficult, to attempt to retrieve the contents of memory and to detect a feature switch. The maintenance-failure hypothesis assumes that the new combination of features overwrites memory, storing these new colour–location combinations in working memory.

As in Experiment 1, four items of different colours rotated around a fixation point, disappearing and reappearing behind an occluding surface, and observers were required to remember which colour appeared in which location while performing a verbal suppression task. At the end of each trial,

all the items stopped while hidden behind the occluders and observers answered the cued-recall question (reporting the location corresponding to the colour cue, or the colour corresponding to the location cue, in separate blocks of trials). Then observers reported whether or not they detected a switch on that trial. When there was a switch, the new colour–location combinations were only seen once for 500 ms (a single eighth of a revolution), so any memory updating must happen within a single interval.

Of principal interest was what colour–location combination observers would report when they missed a feature switch. A tendency to report the old colours in their old locations (the feature combination prior to the switch) would be consistent with the retrieval-failure hypothesis, as success-ful retrieval of the old features suggests that these features were maintained despite missing the switch. A tendency to report the new colours in their new locations (the feature combination after the switch) would be consistent with the maintenance-failure hypothesis, as the original bindings were not maintained. Finally, an equal probability of reporting the new colours in their new locations, the old colours in their old locations, and the other colour–location combinations that were not involved in the switch would be consistent with an encoding-failure hypothesis (not binding the colours to locations before or after the switch). We found this encoding-failure hypothesis unlikely based on the results of Experiment 1, but included it here for the sake of completeness.

Method

Eight new observers participated in the experiment. The method was the same as Experiment 1 except that observers performed both the feature-switch detection task and a cued-recall task on the same trial, with colour-cue and location-cue trials in separate blocks. Observers completed 24 practice and 144 test trials in each of the cued-recall conditions, with the order of conditions counterbalanced across subjects (colour, location; or location, colour).

Results

On average, observers missed 25% ($SEM = 8\%$) of feature switches. The remaining analyses focused on the distribution of cued-recall responses for these miss trials. Responses were categorized as either new-feature (the postswitch colour in the postswitch location), old-feature (the preswitch colour in the preswitch location), or other-features (averaged over the two colour–location combinations that were not involved in the switch). The results show that in both the colour-cue condition and the location-cue condition,

observers reported the new-feature on the vast majority of trials. Figure 3a illustrates that in the colour-cue condition, observers reported the new-feature significantly more than the old-feature, $t(7) = 6.8$, $p < .001$, $r^2 = .87$, or the other-features, $t(7) = 9.4$, $p < .001$, $r^2 = .93$, but there was no significant difference between reports of the old-feature and the other-features ($t < 1$). Similarly, Figure 3b illustrates that in the location-cue condition, observers reported the new-feature significantly more than the old-feature, $t(7) = 6.3$, $p < .001$, $r^2 = .85$, or the other-features, $t(7) = 12.2$, $p < .001$, $r^2 = .96$, but there was no significant difference between reports of the old-feature and the other-feature ($t < 1$).

Discussion

The results indicate that when observers miss feature switches, they report the new colours in their new locations in both cued-recall tasks. Thus, it appears that observers have updated representations of the colour–location combinations, even though they only saw the new feature combinations for 500 ms (one-eighth of a revolution). These results rule out a broad class of explanations for the cause of feature-switch detection errors. Specifically, any
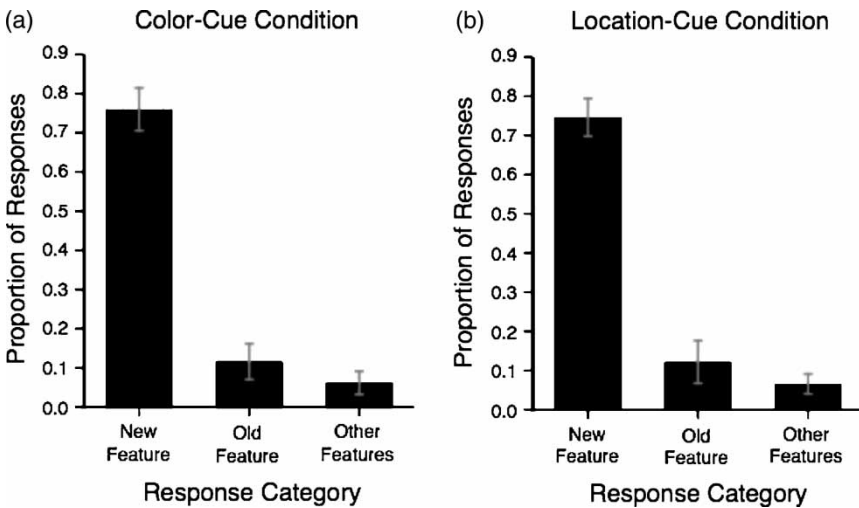

Figure 3. Results of Experiment 2, in terms of the proportion of cued-recall responses in each response category when a feature switch was missed. New-feature corresponds to the new colours in their new locations (the postswitch colours), old-feature corresponds to the old colours in their old locations (the preswitch colours), and other-features corresponds to the other features (those not involved in the feature-switch). (a) The response distribution on colour-cue trials. (b) The response distribution on location-cue trials. Error bars represent one within-subject standard error of the mean, calculated with Cousineau's (2005) modification of Loftus and Masson's (1994) method.

hypothesis that assumes there is not enough time to encode the new feature combinations, or that memory is simply not updated continuously in these dynamic displays, are ruled out by these results. This is consistent with previous research which has shown that colour information can be encoded into visual working memory quite rapidly, on the order of approximately 50 ms per item (Vogel, Woodman, & Luck, 2006). It is possible that stimuli that are encoded this rapidly are particularly susceptible to overwriting and rebinding. The present results are also inconsistent with the hypothesis that feature-switch detection errors are caused by a failure to attempt memory retrieval, as has been shown in previous change-blindness studies (Mitroff et al., 2004; Simons et al., 2002). Although there was a small (nonsignificant) trend for observers to report the old feature combinations more than the other, nonswitch features, the vast majority of cued-recall responses were reports of the new colours in their new locations. Thus, the results are most consistent with the hypothesis that feature-switch detection is difficult because the contents of memory are automatically rewritten and rebound when they are perceived in a new combination. In other words, feature-switch detection errors appear to result from a failure to maintain the initial bindings across the switch.

An alternative possibility is that the initial bindings are maintained, but that there is a bias towards reporting the new feature bindings. On this view, if observers were pressed to report a second guess, the second guess would be more likely to be of the old features (the preswitch colour and location) than the other features (those not involved with the switch). Although it is likely there will be a small trend in this direction, it is unlikely that observers would have maintained the old bindings for a large number of items. Overall accuracy in the cued-recall tasks was nearly the same when there was not a feature switch ($M = 81\%$, $SEM = 4\%$) as when there was a feature switch ($M = 84\%$, $SEM = 4\%$). If the level of accuracy without a feature switch reflects the "full capacity" of memory to store colour–location combinations, then it appears that memory is filled to capacity with the new colour–location combinations after a feature switch (because performance remains at a comparable level). Thus, although there may be some lingering information about the old features, it appears that most of the old information has been displaced by the new feature combinations.

These findings are consistent with previous research, which suggests that bound representations are particularly susceptible to interference from the encoding of new perceptual information (Allen et al., 2006; Wheeler & Treisman, 2002). However, the current results extend these findings by demonstrating that the observed interference is due specifically to an overwriting/rebinding, rather than a more general disruption of stored binding information.

GENERAL DISCUSSION

The current study investigated the ability to encode and maintain colour–location combinations in visual working memory. In Experiment 1, we tested memory for colour–location combinations using two cued-recall tasks (Gajeweki & Brockmole, 2006), and a feature-switch detection task (Saiki, 2002; Wheeler & Treisman, 2002). The results showed that, relative to cued-recall, feature-switch detection underestimates the number of colour–location combinations that can be remembered. These findings indicate that errors in feature switch detection are not due to a failure to encode colour–location bindings into visual working memory (Saiki, 2002, 2003b), but rather indicate a failure at a later stage of working memory processing.

Experiment 2 combined the feature-switch detection task with cued-recall task to determine the nature of feature-switch detection errors by examining which colour–location combinations were remembered when feature-switch detection failed. Observers did not reliably report the preswitch feature combinations (the old colours in their old locations), contradicting the hypothesis that feature-switch errors occur due to a retrieval failure. Instead, observers overwhelmingly reported the postswitch features (the new colours in their new locations), indicating that memory for colour–location combinations had been updated. Thus, it appears that feature-switch detection errors are caused by a failure to maintain the initially bound representations, probably due to the automatic rewriting and rebinding of memory contents in the face of new perceptual input.

Combined, these results clearly indicate that different methods of probing the contents of visual working memory reveal different aspects of feature binding and feature storage. If the goal is to determine the maximum binding capacity of visual working memory (the number of feature combinations that can be remembered), then cued-recall tasks should be preferred over feature-switch detection. Feature-switch detection introduces an additional load at subsequent stages of visual working memory processing, and thus obscures the maximum encoding capacity of visual working memory.

Although feature-switch detection is not ideal for estimating the binding capacity of visual working memory, the difficulty of feature-switch detection is intriguing in its own right, and is worth exploring further. How is it possible to encode the new colour combinations without detecting feature switches? One possibility is that the feature-switch detection task has a higher capacity demand than the cued-recall tasks. For example, one way to detect a feature switch would be to store all of the preswitch features and all of the postswitch features, and then compare them. This would require the storage of eight total colour–location combinations in the current experiments, which clearly exceeds the estimated upper bounds of visual working

memory capacity, which appears to be about four items (Alvarez & Cavanagh, 2004; Luck & Vogel, 1997). In contrast, the cued-recall tasks could be performed by only remembering four colour–location combinations at a time. Another possibility is that encoding information into visual working memory is an automatic process (i.e., does not require active cognitive control), but that the detection of recombinations of features is a less natural task that requires active cognitive control. This additional processing demand could increase the likelihood of errors in the feature-switch detection task. Future work will be required to explore these and other possible explanations for the relative difficulty of maintaining bound object representations when a feature switch occurs.

CONCLUSION

The results indicate that feature-switch detection tasks will underestimate the number of bound objects that can be encoded into memory relative to cued-recall tasks, suggesting that cued-recall is the preferred method for determining the maximal binding capacity of visual working memory. However, the feature-switch detection task appears to strain processes involved with maintaining the contents of working memory (due to rewriting and rebinding of memory contents by new perceptual input), making this task valuable for gaining insight into the fragile nature of working memory for bound object representations. These results could have important implications for understanding the nature of storage and access to bound object representations in visual working memory.

REFERENCES

Allen, R. J., Baddeley, A. D., & Hitch, G. J. (2006). Is the binding of visual features in working memory resource-demanding? *Journal of Experimental Psychology: General*, *135*(2), 298–313.

Alvarez, G. A., & Cavanagh, P. (2004). The capacity of visual short-term memory is set both by visual information load and by number of objects. *Psychological Science*, *15*(2), 106–111.

Atkinson, R. C., & Shiffrin, R. M. (1968). Human memory: A proposed system and its control processes. In K. W. Spence & J. T. Spence (Eds.), *The psychology of learning and motivation: Advances in research and theory* Vol. 2 (pp. 89–195). New York: Academic Press.

Brainard, D. H. (1997). The Psychophysics Toolbox. *Spatial Vision*, *10*, 433–436.

Cousineau, D. (2005). Confidence intervals in within-subject designs: A simpler solution to Loftus and Masson's method. *Tutorial in Quantitative Methods for Psychology*, *1*(1), 42–45.

Cowan, N. (2001). The magical number 4 in short-term memory: A reconsideration of mental storage capacity. *Behavioral and Brain Sciences*, *24*(1), 87–114; discussion 114–185.

Cowey, A. (1985). Aspects of cortical organization related to selective attention and selective impairments of visual attention. In M. I. Posner & O. S. M. Marin (Eds.), *Attention and*

performance XI: Attention and neuropsychology (pp. 41–62). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Gajewski, D. A., & Brockmole, J. R. (2006). Feature bindings endure without attention: Evidence from an explicit recall task. *Psychonomic Bulletin and Review*, *13*(4), 581–587.

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

Luck, S. J., & Vogel, E. K. (1997). The capacity of visual working memory for features and conjunctions. *Nature*, *390*(6657), 279–281.

Mitroff, S. R., Simons, D. J., & Levin, D. T. (2004). Nothing compares 2 views: Change blindness can occur despite preserved access to the changed information. *Perception and Psychophysics*, *66*(8), 1268–1281.

Pashler, H. (1988). Familiarity and visual change detection. *Perception and Psychophysics*, *44*(4), 369–378.

Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: Transforming numbers into movies. *Spatial Vision*, *10*, 437–442.

Saiki, J. (2002). Multiple-object permanence tracking: Limitation in maintenance and transformation of perceptual objects. *Progress in Brain Research*, *140*, 133–148.

Saiki, J. (2003a). Feature binding in object-file representations of multiple moving items. *Journal of Vision*, *3*(1), 6–21.

Saiki, J. (2003b). Spatiotemporal characteristics of dynamic feature binding in visual working memory. *Vision Research*, *43*(20), 2107–2123.

Saiki, J., & Miyatsuji, H. (2007). Feature binding in visual working memory evaluated by type identification paradigm. *Cognition*, *102*(1), 49–83.

Simons, D. J. (2000). Current approaches to change blindness. *Visual Cognition*, *8*(1/2/3), 1–15.

Simons, D. J., Chabris, C. F., Schnur, T., & Levin, D. T. (2002). Evidence for preserved representations in change blindness. *Consciousness and Cognition*, *11*(1), 78–97.

Smith, J. E. K. (1982). Simple algorithms for M-alternative forced choice calculations. *Perception and Psychophysics*, *31*(1), 95–96.

Stefurak, D. L., & Boynton, R. M. (1986). Independence of memory for categorically different colors and shapes. *Perception and Psychophysics*, *39*(3), 164–174.

Treisman, A. (1996). The binding problem. *Current Opinion in Neurobiology*, *6*(2), 171–178.

Treisman, A. (1998). Feature binding, attention and object perception. *Philosophical Transactions of the Royal Society of London*, *353*(1373), 1295–1306.

Treisman, A., & Schmidt, H. (1982). Illusory conjunctions in the perception of objects. *Cognitive Psychology*, *14*(1), 107–141.

Treisman, A., & Zhang, W. (2006). Location and binding in visual working memory. *Memory and Cognition*, *34*(8), 1704–1719.

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*(1), 92–114.

Vogel, E. K., Woodman, G. F., & Luck, S. J. (2006). The time course of consolidation in visual working memory. *Journal of Experimental Psychology: Human Perception and Performance*, *32*(6), 1436–1451.

Wheeler, M. E., & Treisman, A. M. (2002). Binding in short-term visual memory. *Journal of Experimental Psychology: General*, *131*(1), 48–64.

Wolfe, J. M., Klempen, N., & Dahlen, K. (2000). Postattentive vision. *Journal of Experimental Psychology: Human Perception and Performance*, *26*(2), 693–716.

Zeki, S. (1993). *A vision of the brain*. Oxford, UK: Blackwell.