

Integrating Sequential Arrays in Visual Short-Term Memory

Yuhong Jiang, Arjun Kumar, and Timothy J. Vickery

Harvard University, Cambridge, MA, USA

Abstract. Are sequential visual arrays represented as separate images or as a combined image in visual short-term memory (VSTM)? Proponents of the integration account suggest that an image of the first array is gradually formed and integrated with an image of the second to produce a combined representation. This view is evidenced by successful performance in an empty-cell detection task. In this task, on a 4×4 square matrix, 7 locations are occupied on a first array, followed by a variable interval, and then by 8 other occupied locations on a second array. Subjects' success in identifying the remaining empty cell has been taken as evidence for integration. In this study, we show that success in this task can be better accounted for by a convert-and-compare process than by an integration process. We conclude that VSTM only supports limited integration across sequential arrays.

Keywords: visual short-term memory, visual integration

Introduction

Visual short-term memory (VSTM) allows a visual percept to be retained momentarily after its disappearance (Irwin, 1991; Pashler, 1988; Phillips, 1974). It creates temporal continuity in a constantly changing visual environment, yet it has severe capacity limits. Only four objects or six spatial locations can be retained in VSTM (Jiang, Olson, & Chun, 2000; Luck & Vogel, 1997; Pashler, 1998; Phillips, 1974). Research effort in the past decade has focused primarily on the representation of a single visual display. Yet, in many everyday activities, visual information not only occupies space but also evolves over time. Driving requires the constant updating of information gathered from many instances. When crossing a busy street we look left and then right. Even the simplest social interactions depend on holding information across space and time. Successive visual input may interact by competing for space in a limited capacity VSTM (Jiang & Kumar, in press; Phillips & Christie, 1977), producing proactive or retroactive interference (Jiang, 2004), or becoming integrated into a combined display (Brockmole, Wang, & Irwin, 2002). For this reason, we cannot understand the real-world representation of sequential displays in VSTM simply by investigating how an isolated temporal instance is remembered. This study examines how good humans are at integrating sequential displays in VSTM.

Recent studies have revealed that, despite its limited capacity, VSTM is capable of integrating information gathered across sequential arrays (Brockmole, Irwin, & Wang, 2003; Brockmole, et al., 2002; Brockmole & Wang, 2003). This is shown in an empty-cell detection task developed initially by Eriksen and Collins (1967) and Di Lollo (1980). In this task, subjects are presented with two dot arrays in a 5×5 square matrix. The first array occupies 12 randomly selected locations in the matrix, and the second array occupies another 12 locations, such that the two arrays, if superimposed, would leave one cell empty. The two arrays are presented briefly and separated by a variable stimulus-onset-asynchrony (SOA). Di Lollo (1980) found that empty-cell detection is successful within an SOA of about 100 ms, but it deteriorates rapidly as the SOA increases from 100 ms to 500 ms. This suggests that visual integration is good within iconic memory, but poor within visual short-term memory.

However, as one further increases the SOA from 500 ms to 2,500 ms, performance gradually improves until it reaches an asymptote at an SOA of approximately 1,300 ms (Brockmole et al., 2002). This suggests that visual integration in VSTM is possible, provided that the SOA is relatively long. Brockmole et al. propose that the long SOA is necessary for subjects to form a mental image of the first array so that, when the second array appears, it can be integrated with this

image. This hypothesis will be referred to as the *image-percept integration hypothesis*. We will use “image-formation” to describe the process of forming a mental representation of the first array. Based on the level of performance, Brockmole et al. estimate that in a 4×4 square matrix, about 6.2 dots from the first array and 7.8 dots from the second array can be retained. The estimated capacity is even higher when subjects are tested using a 5×5 square matrix, at about 9.6 dots from the first array and 11.7 dots from the second array. Such estimated capacity is several times larger than that estimated from other studies (4 objects or 6 spatial locations; Jiang et al., 2000; Luck & Vogel, 1997; Wheeler & Treisman, 2002), raising the question of how subjects expand their capacity.

In this article, we challenge the view that the human VSTM is efficient at integrating information from sequential arrays. We propose, instead, that although VSTM supports effective comparison between simple representations, it does not support the *creation* of an entirely new representation on the basis of *two* separate images. This account further proposes that lacking the ability to integrate, subjects adopt a convert-and-compare strategy to succeed at the empty-cell detection task, by converting the first array into its negative (i.e., remembering only the unoccupied locations in the first array and then comparing this negative image with the second array to spot the difference. Whichever empty location from the first array is not subsequently filled by a dot in the second would be the missing cell). As the SOA increases, subjects’ ability to extract and retain the empty locations on the first array improves, accounting for the increased accuracy in finding the empty cell. We will refer to this process as “negative conversion,” and to our hypothesis as the *limited-integration hypothesis*.

The two accounts differ in their assumptions about how VSTM is limited. The image-percept integration account postulates that VSTM has a high capacity of about 14–20 locations, and that it supports integration across sequential arrays. The limited-integration account postulates that VSTM has a small capacity of about 6 spatial locations, and that it does not effectively support the creation of a combined representation obtained by integrating two images. Our study is designed to test these competing hypotheses.

Experiment

We approach this question by comparing performance in the empty-cell detection task with that in the change detection task. Unlike empty-cell detection,

change detection (Rensink, 2002) does not require integration. In this task, subjects are shown 7 locations on array 1, and then these locations plus a new one on array 2. Subjects are asked to detect how the first array has changed by reporting the new location. This task is essentially the same as that used by Phillips (1974) and Irwin (1991), and is now routinely used to measure VSTM capacity. To detect a change across two arrays, one simply needs to form an image of array 1 and compare it with array 2. Thus, forming an image of array 1 (image-formation) fits the requirement of change detection.

In contrast, image-formation fits empty-cell detection only according to the image-percept integration account. The alternative view, the limited-integration account, predicts that subjects rely on negative conversion to find the empty-cell. If subjects extract and retain the negative of array 1, they will gradually lose the positive, so executing the change detection task via negative conversion would be difficult. As an analogy, consider Rubin’s Face-Vase figure that leads to bistable percepts. When people perceive the faces in the figure, they have difficulty perceiving the vase simultaneously, and vice versa. In short, the image-percept integration account predicts that a single process, image-formation, is beneficial for empty-cell detection as well as for change detection, whereas the limited-integration account predicts that the suitable process for change detection (image-formation) is unsuitable for empty-cell detection, and vice versa.

To test the above predictions, we randomly intermix empty-cell detection and change-detection tasks within a block, in varying proportions. In both conditions, subjects are presented with 7 filled locations on array 1 and 8 filled locations on array 2. The two arrays are separated by a variable inter-stimulus-interval (ISI) ranging from 200 ms to 1,500 ms. In the *75% change-detection* condition, 75% of trials require change detection while the others require empty-cell detection. Subjects are informed of this proportion at the start of the block and are asked to optimize accuracy for change detection because this is in the majority. In the *75% empty-cell detection* condition, 75% of trials require empty-cell detection, which is emphasized as the primary task.

As subjects try to optimize the primary task, the other task can be considered as a probe for how the primary task is accomplished. Because the image-percept integration account postulates that image-formation fits empty-cell detection as well as change detection, it predicts that (1) whether a given task is in the majority or minority should not influence the effect of ISI, and (2) the effect of ISI should be similar for change detection and for empty-cell detection. In con-

trast, the limited integration account postulates that image-formation fits change detection, but negative-conversion fits empty-cell detection. If change detection is in the majority, subjects will not convert array 1 into its negative, so empty-cell detection will not benefit from a long ISI. Conversely, if empty-cell detection is in the majority, subjects will convert and retain the negative of array 1. As they become more successful, their empty-cell detection will improve, but their change detection will deteriorate. Figure 1 shows a sample display of the two conditions.

Method

Participants

Eight college students (19–28 years old) participated for payment.

Materials and Procedure

On each trial, subjects were presented with a 4×4 square matrix that subtended $12.8^\circ \times 12.8^\circ$. The color of the background was gray and the color of the grid was blue. The first array contained 7 filled black dots (each 3.2°) presented at randomly selected cells of the

matrix. The array was presented for 2 refresh cycles (27 ms) and erased. After a variable inter-stimulus-interval (ISI), a second array of 8 filled black dots (each 3.2°) was presented. The second array was presented for 2 refresh cycles (27 ms) and erased. A small question mark was then presented at the center of the empty grid as a prompt for subjects to make their decisions. Subjects were asked to decide which cell contained the target and press the spacebar once they had made their decision. An array of 16 letters was then brought up, and subjects were asked to type in the letter at the target position.

There were two possible tasks: change detection and empty-cell detection. Which task was relevant was indicated by the similarity between array 1 and array 2. In the *change detection* task, 7 of the 8 locations on array 2 were the same 7 locations as on array 1, while the 8th location was a new location. Subjects were asked to report the position of the difference if the two arrays looked similar. In the empty-cell detection task, all 8 locations on array 2 were different from the 7 locations on array 1, which left 1 cell unoccupied by either array. Subjects were asked to report the position of the empty cell if the two arrays looked different and complimentary. Pilot studies showed that the difference between the two types of trials was salient and that the instructions were easy to follow.

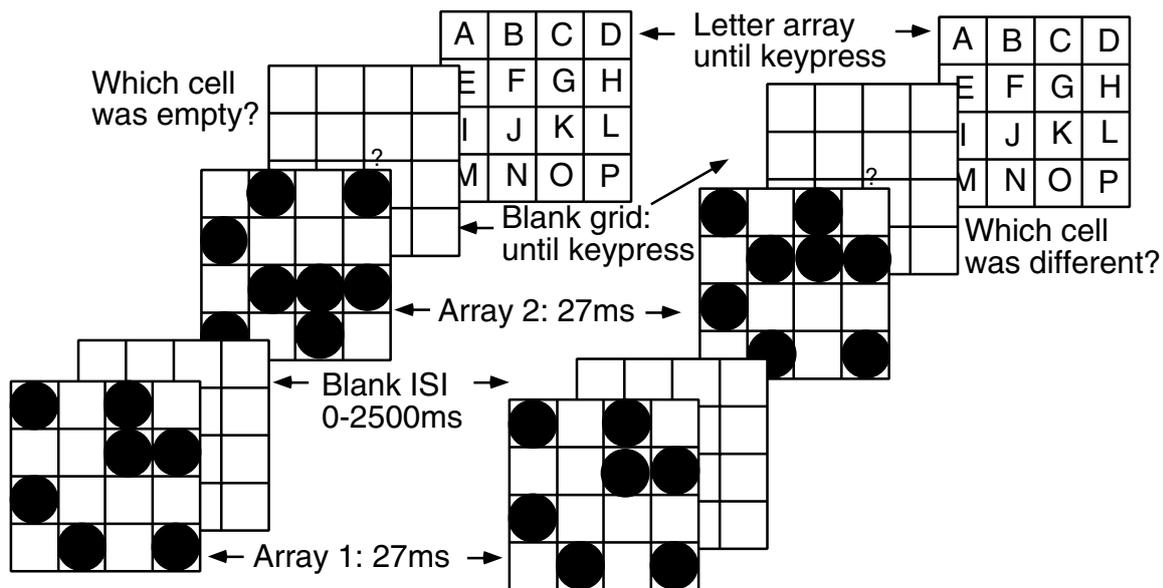


Figure 1. Sample presentation sequences tested in the experiment. The *empty-cell detection* task is shown on the left and the *change detection* task is shown on the right. Subjects were asked to find the empty cell if the two arrays appeared complimentary and to find the difference if the two arrays were similar. Upon their decision they pressed the spacebar and typed in the letter at the target position. The two conditions were randomly intermixed, so that subjects could not predict the task until the second array had been presented.

Design

There were 2 baseline blocks and 4 experimental blocks. The two baseline blocks were: *100% change detection* and *100% empty-cell detection*. Each block contained 10 practice and 96 experimental trials. There were 32 trials in each of the three ISIs (200, 500, or 1,500 ms), presented in a randomly intermixed order.

There were four experimental blocks: two were *75% change detection* and the other two *75% empty-cell detection*. Each block contained 5 practice and 96 experimental trials. Of the 96 trials, 75% (72 trials) belonged to the majority condition and the other 25% (24 trials) to the minority condition. Three ISIs (200, 500, or 1,500 ms) were tested. The six blocks were presented in a randomly determined order. Prior to the presentation of the block, subjects were informed of the proportion of trials falling into each condition. They were asked to optimize their overall accuracy.

Results

We plot the percentage of correct and incorrect trials, separately for change detection and empty-cell detection tasks, and separately for different conditions. Figure 2 shows the results.

1. Baseline Blocks: 100% Conditions

100% Change Detection

Accuracy showed a significant decrease as the ISI increased from 200, 500, to 1,500 ms, $F(2, 14) = 8.18$, $p < .004$. This suggests a certain degree of memory decay over a period of 1,500 ms, consistent with a pattern observed by Phillips (1974) and Irwin (1991). Memory accuracy dropped by 11.1%, which was produced largely by subjects' increasing errors in misidentifying one of the repeated locations as the difference (a 10% increase), $F(2, 14) = 5.50$, $p < .017$.

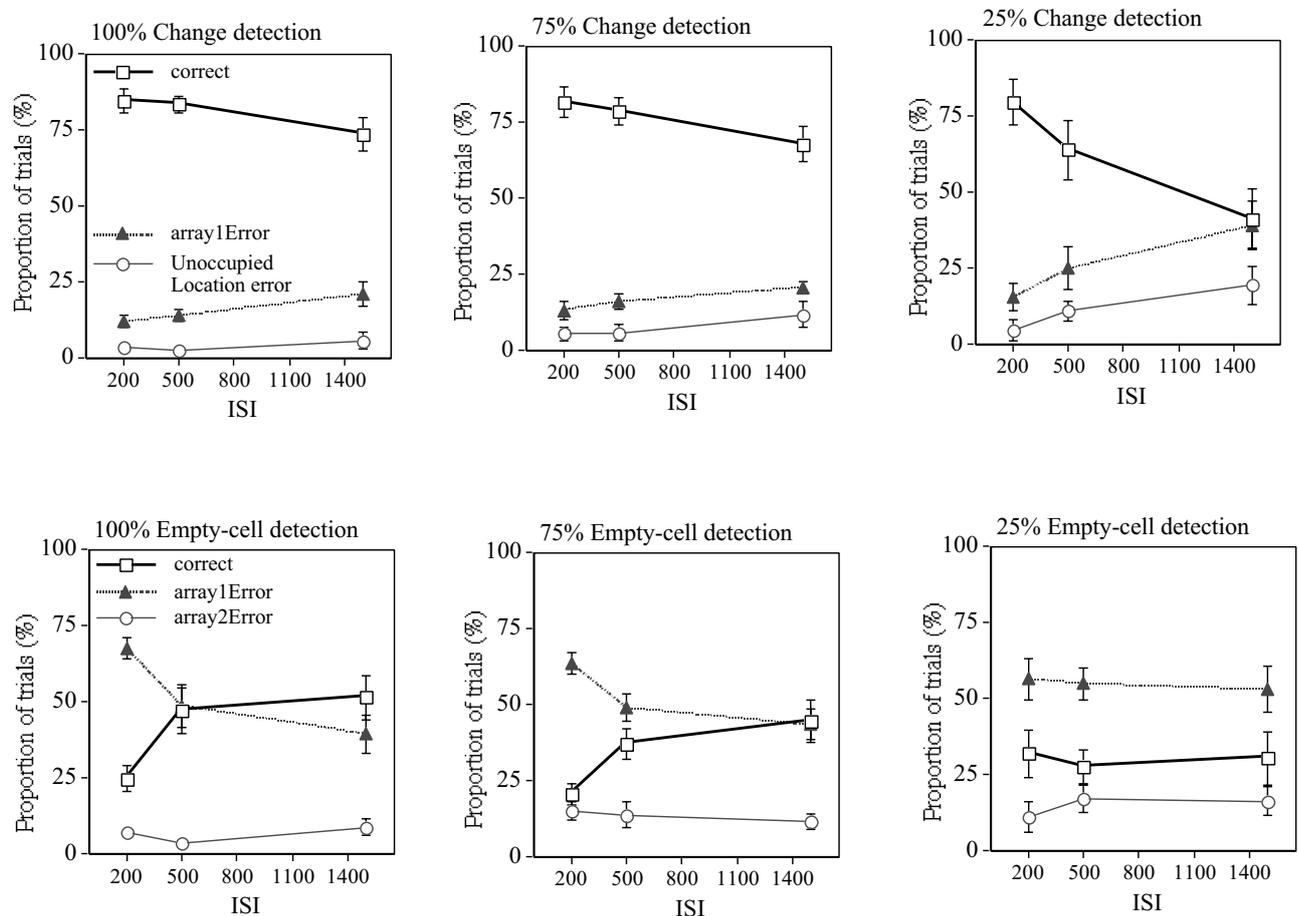


Figure 2. Results from the experiment. The top row shows performance in the change detection task; the bottom row shows performance in the empty-cell detection task.

100% Empty-Cell Detection

Accuracy showed a significant improvement as the ISI increased from 200, 500, to 1,500 ms, $F(2, 14) = 13.15$, $p < .001$. Memory accuracy increased by 26.7%, which was largely accounted for by the reduction in array 1 errors (a 27% decrease, $p < .001$).

2. Experimental Blocks: 75% Conditions

75% Change Detection

(1) Performance in the change-detection task (75%)

Here, too, accuracy showed a significant decrease as the ISI increased, $F(2, 14) = 29.50$, $p < .001$. Memory accuracy dropped by approximately 14%.

Change Detection: 100% vs. 75%. A direct comparison between the 100% and 75% blocks for change detection revealed a significant main effect of block, $F(1, 7) = 6.68$, $p < .036$, showing better memory (by 3%) when all trials required change detection; a significant main effect of ISI, $F(2, 14) = 22.66$, $p < .001$, showing a significant memory decay. The interaction between the two factors was not significant, $F(2, 14) < 1$, suggesting that the degree of memory drop was comparable.

(2) Performance in the Empty-Cell Detection Task (25%)

Accuracy to detect the empty-cell was not significantly influenced by ISI, when the empty-cell detection task was in the minority of the trials, $F(2, 14) < 1$, *ns*. Accuracy stayed at approximately 30% across the three ISIs.

Empty-Cell Detection: 100% vs. 25%. An ANOVA on proportion (100% vs. 25%) and ISI revealed significant main effects of proportion, $F(1, 7) = 6.39$, $p < .039$, ISI, $F(2, 14) = 5.99$, $p < .013$; and a significant interaction, $F(2, 14) = 6.55$, $p < .01$. Thus, compared with the baseline 100% condition, when the empty-cell detection task was in the minority, accuracy was lower and showed no improvement across time.

75% Empty-Cell Detection

(1) Performance in the Change-Detection Task (25%)

When change detection was in the minority, accuracy plummeted as the ISI increased, $F(2, 14) = 8.56$, $p <$

.004. Accuracy was 80% when ISI was 200 ms and decreased to 41% when ISI was 1,500 ms, a 39% drop.

Change Detection: 100% vs. 25%. An ANOVA on proportion (100% vs. 25%) and ISI showed significant main effects of proportion, $F(1, 7) = 13.24$, $p < .008$; ISI, $F(2, 14) = 11.78$, $p < .001$, and their interaction, $F(2, 14) = 4.38$, $p < .033$. Thus, compared with the baseline 100% condition, when the change detection task was in the minority, accuracy was lower and showed a much more dramatic drop across time.

(2) Performance in the Empty-Cell Detection Task (75%)

Accuracy showed a significant improvement as the ISI increased, $F(2, 14) = 15.03$, $p < .001$, of about 24%, which was largely accounted for by a 20% decrease in array 1 error.

Empty-Cell Detection: 100% vs. 75%. This analysis showed a significant main effect of ISI, $F(2, 14) = 21.58$, $p < .001$. The main effect of proportion failed to reach significance, $F(1, 7) = 2.25$, $p > .15$. The interaction effect was also not significant, $F(2, 14) < 1$, *ns*.

Discussion

By directly manipulating the proportion of trials in the change detection and the empty-cell detection conditions, we strengthened the idea that image-formation could not account for how empty-cell detection was carried out. When empty-cell detection was tested in a single block, or when it was the majority of trials, accuracy to detect the empty cell improved as the ISI increased. Yet, when change detection dominated the block, biasing subjects to extract and form a positive image of array 1, empty-cell detection showed no improvement as the ISI increased. In addition, accuracy in the change detection task showed a decay of about 10–15% when this task was in the majority, but the reduction in accuracy precipitated to nearly 40% when it was in the minority. These data suggest that the best strategy for change detection is detrimental to empty-cell detection, and that the best strategy for empty-cell detection is detrimental to change detection. This pattern of results fits better with the limited integration account than with the image-percept integration account.

The decrease in performance in the 25% conditions may partly be accounted for by the change in task set, from the majority task to the minority task (we thank Dr. James Brockmole for raising this possibility). However, changes in task set and their result-

ing confusion should have a generally detrimental effect on all trials of the minority task, independent of the ISI. But this was not what we found. In particular, when the ISI was 200 ms, performance in the minority task was comparable to performance in the 100% baseline condition of that task (p values $> .20$). Given that accuracy did not decrease in general, the rise and fall of performance at later ISIs cannot be attributed entirely to confusion and task-set switching.

How are two sequential visual arrays represented in visual short-term memory (VSTM)? Are they effectively integrated, as suggested by the image-percept integration account (Brockmole et al., 2002), or not, as suggested by the limited integration account? Our study has provided evidence for the latter. The following is a summary.

First, if one assumes that subjects detect the empty cell by integrating across the two arrays, then one would have to accept that as many as 14–20 spatial locations could be retained in VSTM; otherwise subjects would not have reached an accuracy of 50% or higher. The assumption of high-capacity is inconsistent with many other studies that have revealed a low-capacity of VSTM. Several studies testing spatial VSTM have observed that subjects are able to retain approximately six spatial locations but not more. The exact number of locations retained depends on whether subjects are allowed to “chunk” individual items into a global pattern (Jiang et al., 2000). When forced to retain individual items independent of one another, memory accuracy is quite poor (e.g., 70% on a 2AFC task) even when only three locations have to be retained; when allowed to retain the global pattern, approximately six locations can be retained. Still, even with configural processing, the capacity of VSTM for spatial locations is much smaller than 14 locations. The image-percept integration hypothesis would have to account for the discrepancy in its estimated capacity and that found by many others (e.g., Jiang et al., 2000; Luck & Vogel, 1997).

Second, the ISI for asymptote performance is approximately 1,300 ms. According to the image-formation hypothesis (see also the consolidation hypothesis, Jiang, 2004), such a long time window is necessary for one to form an image of array 1. Yet, other studies have suggested that visual processing takes much shorter: a single glimpse of 200 ms is sufficient for us to detect the presence of animals in masked, complex images (Thorpe, Fixe, & Marlot, 1996), for face recognition (Liu, Harris, & Kanwisher, 2002), and for scene perception (Potter, 1976).

Alternatively, one might argue that such a long time window is necessary for “protecting” image 1 from any possible disruption by array 2 (Jiang, 2004).

However, sensory masking lasts about 100 ms (Sperling, 1960), conceptual masking lasts about 300 ms (Potter, 1976), and the attentional dwell-time (or consolidation time) takes about 500 ms (Duncan, Ward, & Shapiro, 1994). The puzzle is why 1,300 ms is necessary for optimal performance in empty-cell detection. An optimal ISI of 1,300 ms cannot be accounted for by the gradual formation of a combined visual image, because the integration between images 1 and 2 cannot start until *after* the presentation of array 2, so one cannot start forming the combined image during the SOA. One might propose that a window for potential masking is lengthened by an anticipated need to integrate, i.e., integration produces a retrograde effect that lengthens the window for image formation. Yet, this fails to account for the difference between empty-cell detection and change detection, as discussed below.

Third, if subjects become increasingly successful at extracting image 1, then image-formation should be beneficial to empty-cell detection as well as to change detection, which relies on retaining array 1 and comparing it with array 2. However, performance in these two tasks does not rise and fall together: they change in opposite directions as the ISI increases. If a block contains mostly change detection trials, then empty-cell detection does not improve over time; if a block contains mostly empty-cell detection trials, then as empty-cell detection improves, change detection deteriorates precipitately. Thus, whatever is extracted during change detection is incompatible with whatever is formed during empty-cell detection. The only possibility that this would happen is if the two tasks rely on complementary information.

In comparison, a convert-and-compare strategy fits a wider range of data. We do not have to hypothesize a high-capacity system, nor do we assume effective integration. Suppose one can retain 4 empty locations during the SOA – a modest estimate that is slightly worse than the typical estimation – then accuracy to compare these locations with eight new locations would reach 50%. Note that subjects rarely commit array 2 errors, perhaps because when uncertain, they would guess from locations not filled on array 2.

This alternative account can easily explain why the optimal ISI was about 1,300 ms. This is because during the ISI a new image – the negative – needs to be created. This process can presumably take 1,300 ms, even though forming a positive image takes much shorter. Finally, a convert-and-compare strategy is incompatible with image formation, explaining the discrepancy between change detection and empty-cell detection.

A recent study has provided apparently conflicting conclusions with respect to whether subjects relied on

“negative-conversion” in the empty-cell detection task (Brockmole et al., 2003). The logic of the study is as follows: If subjects gradually form an image of array 1, then as the ISI increases and their image-formation is more successful, subjects’ *attention* should be allocated to the positive of array 1. In turn, if a probe item falls on the positive locations, probe discrimination should be faster than if a probe item falls on an empty location.

To test whether attention was focused on the positive locations or on the empty locations, Brockmole et al. (2003) presented 5 dots on array 1 and 10 on array 2. On 1/3 of trials, the second array was not presented. Instead, after a variable ISI, a probe fell on a previously filled location 50% of time (“ON”), and on one of the empty locations 50% of time (“OFF”). RT to identify the probe was significantly faster for ON than for OFF, when the ISI was 1,450–2,350 ms, but not when the ISI was 100–1,000 ms, or when it was longer than 2,800 ms. This suggests that attention was allocated to the ON locations at a relatively long ISI, supporting the gradual formation of array 1.

This conclusion, however, is weakened for the following reasons. First, because the probe would fall on the five filled locations 50% of time and the 11 empty locations 50% of time, the probability that the probe would fall on *any* of the filled locations was twice as likely as on *any* of the unfilled locations. The uneven probability might have induced subjects to anticipate a probe at the filled locations, producing an ON-advantage (we thank Dr. Andrew Hollingworth for making this observation). Second, the probe effect was observed only when the ISI was 1,450–2,350 ms, not when it was 1,000 ms or 2,800 ms. Because accuracy to detect the empty-cell was high at ISIs of 1,000 ms and 2,800 ms, there was no reason why a probe effect should not also be observed at these lags. Finally, in an ongoing study, we were unable to replicate the probe-detection results using conditions as similar as possible to those used by Brockmole et al. (2003). We suspect that the probe task used in Brockmole et al. (2003) and in our failure-to-replicate was not sensitive enough to pinpoint which locations were held in VSTM, because the filled and empty locations were highly intertwined in the same general region. The change detection task used here was a more sensitive probe task that tapped directly into the content of VSTM. For these reasons, we do not think there is strong evidence against our current view. Nonetheless, future studies are necessary to resolve the discrepancy between different kinds of probe tasks.

While identifying negative-conversion as a detour through which subjects succeed at the empty-cell detection task, we note that it is highly unlikely that peo-

ple actually perform negative conversion in everyday vision. This strategy is seen perhaps only in laboratory tasks that necessitate its application. Lacking a good ability to integrate, and finding no motivation to combine two temporal events into one, we are left with two separate images, with the trailing event dominating visual perception and working memory (Jiang & Kumar, in press).

What did we lose by favoring a limited integration hypothesis? For one thing, it is highly unlikely that VSTM lacks any ability to integrate. Instead, we propose that limited integration *can* take place in VSTM, although it does not support effective integration of large set sizes. More critically, by rejecting the possibility that a new representation can be effectively formed in VSTM by pooling information from two arbitrary arrays, we ascribe to VSTM a much more modest role in integration. It is highly unlikely that VSTM plays a major role in creating a sense of stability of the visual environment. Although limited, a system with limited integration in VSTM is perhaps better than a system that integrates information gathered from entirely different arrays: it prioritizes the new display and protects us from incorrectly combining information from totally different temporal moments.

How do people retain multiple visual percepts in everyday activities? We suggest that this is achieved largely by representing events that have spatio-temporal continuity. Tracking multiple objects over time is a powerful way to achieve perceptual continuity (Scholl & Pylyshyn, 1999). However, the presentation of two arbitrary arrays, such as those used in the empty-cell detection task, most likely indicates two different events. Under such conditions, perhaps we do not, and cannot, integrate sequential arrays.

Acknowledgement

This research was supported in part by a Helen Hay Whitney Fellowship and the Clark Fund to Y. J. We thank Sidney Burks for data collection, and James Brockmole, Andrew Hollingworth, Karl Klauer, Mary Potter, and Joo-Hyun Song for comments and suggestions.

References

- Brockmole, J. R., Irwin, D. E., & Wang, R. F. (2003). The locus of spatial attention during the temporal integration of visual memories and visual percepts. *Psychonomic Bulletin & Review*, *10*, 510–515.

- Brockmole, J. R., & Wang, R. F. (2003). Integrating visual images and visual percepts across time and space. *Visual Cognition*, *10*, 853–873.
- Brockmole, J. R., Wang, R. F., & Irwin, D. E. (2002). Temporal integration between visual images and visual percepts. *Journal of Experimental Psychology: Human Perception & Performance*, *28*, 315–334.
- Di Lollo, V. (1980). Temporal integration in visual memory. *Journal of Experimental Psychology: General*, *109*, 75–97.
- Duncan, J., Ward, R., & Shapiro, K. (1994). Direct measurement of attentional dwell time in human vision. *Nature*, *369*, 313–315.
- Eriksen, C. W., & Collins, J. F. (1967). Some temporal characteristics of visual pattern perception. *Journal of Experimental Psychology*, *74*, 476–484.
- Irwin, D. E. (1991). Information integration across saccadic eye movements. *Cognitive Psychology*, *23*, 420–456.
- Jiang, Y. (2004). Time window from visual images to visual short-term memory: Consolidation or integration? *Experimental Psychology*, *51*, 45–51.
- Jiang, Y., Olson, I. R., & Chun, M. M. (2000). Organization of visual short-term memory. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, *26*, 683–702.
- Jiang, Y., & Kumar, A. (in press). Visual short-term memory for two sequential arrays: One integrated representation or two separate representations? *Psychonomic Bulletin & Review*.
- Liu, J., Harris, A., & Kanwisher, N. (2002). Stages of processing in face perception: An MEG study. *Nature Neuroscience*, *5*, 910–916.
- Luck, S. J., & Vogel, E. (1997). The capacity of visual working memory for features and conjunctions. *Nature*, *309*, 279–281.
- Pashler, H. (1988). Familiarity and visual change detection. *Perception & Psychophysics*, *44*, 369–378.
- Phillips, W. A. (1974). On the distinction between sensory storage and short-term visual memory. *Perception & Psychophysics*, *16*, 283–290.
- Phillips, W. A., & Christie, D. F. M. (1977). Interference with visualization. *Quarterly Journal of Experimental Psychology*, *29*, 637–650.
- Potter, M. C. (1976). Short-term conceptual memory for pictures. *Journal of Experimental Psychology*, *2*, 509–522.
- Rensink, R. A. (2002). Change detection. *Annual Review of Psychology*, *53*, 245–277.
- Scholl, B. J., & Pylyshyn, Z. W. (1999). Tracking multiple items through occlusion: Clues to visual objecthood. *Cognitive Psychology*, *38*, 259–290.
- Sperling, G. (1960). The information available in brief visual presentation. *Psychological Monographs: General and Applied*, *74*, 1–29.
- Thorpe, S., Fize, D., & Marlot, C. (1996). Speed of processing in the human visual system. *Nature*, *381*, 520–522.
- Wheeler, M. E., & Treisman, A. M. (2002). Binding in short-term visual memory. *Journal of Experimental Psychology: General*, *131*, 48–64.

Received December 25, 2003

Revision received March 25, 2004

Accepted March 26, 2004

Address for correspondence

Yuhong Jiang
 Department of Psychology
 Harvard University
 33 Kirkland Street, Room 820
 Cambridge, MA 02138
 USA
 Tel. +1 617 496-4486
 Fax +1 617 495-3728
 E-mail yuhong@wjh.harvard.edu
