

# Connecting the past with the present: How do humans match an incoming visual display with visual memory?

Joo-Hyun Song

Departments of Psychology, Harvard University,  
Cambridge, MA, USA



Yuhong Jiang

Departments of Psychology, Harvard University,  
Cambridge, MA, USA



Extensive cognitive research has been devoted to the sensitivity of the visual system to invariant statistical information. For example, many studies have shown that performance improves when a visual display is presented repeatedly. But what allows humans to connect the current visual input to previous memory? Is the connection made only when the entire incoming display matches with a previous memory, or can retrieval rely on an incomplete match between the input and a learned display? Using a visual search task, we show that (1) once a repeated display is learned, subjects can retrieve it even when an incoming display only matches it in 3-4 locations; (2) however, early during learning, repetition of a small proportion of a display is not enough to establish a strong memory trace for the repeated locations. We suggest that the retrieval of a well-established visual memory can proceed even if an incoming display partly matches the previous memory.

Keywords: visual search, contextual cueing, implicit visual learning

## Introduction

The human visual system operates with stunning efficiency: A single glimpse at a complex natural scene is sufficient for detection of the presence of animals and vehicles (Thorpe, Fize, & Marlot, 1996; Li, VanRullen, Koch, & Perona, 2002). Such efficiency relies on at least two functions: object recognition and scene statistical analysis. Extensive research has been devoted to studying both functions. For example, many studies have examined the mechanisms that allow one to recognize an object as a known object. These mechanisms include template matching, feature extraction, and structural description (Biederman, 1987), among others (Palmer, 1999). At the same time, many other studies have tested the visual system's sensitivity to statistical information, particularly visual information that occurs repeatedly in the past. These studies show that humans are extremely efficient at extracting regular, or invariant, visual information that occurs repeatedly. For example, humans are sensitive to repeated spatial layout (Chun & Jiang, 1998), temporal sequence (Nissen & Bullemer, 1987; Olson & Chun, 2001), motion trajectories (Chun & Jiang, 1999), target location (Miller, 1988), and object pairs (Chun & Jiang, 1999; Fiser & Aslin, 2001).

Surprisingly, few studies on visual statistical learning have investigated how the currently encountered information is linked with one's previous visual memory. For example, we navigate our own neighborhood with higher efficiency than when we navigate a novel city, presumably because we rely on past knowledge about a familiar environment. Suppose we then move to another city and revisit our hometown a few years later. Will we continue to pro-

cess the old spatial layout with superb efficiency? If so, does this efficiency require the preservation of the entire layout of our old neighborhood, or can we tolerate mismatches produced by new changes in the current layout?

The present study relies on a paradigm known as "contextual cueing" to address these questions. In the following sections, we shall first review relevant literature on contextual cueing and then present three experiments that address the retrieval of a well-learned visual layout.

## Contextual cueing

To examine human's proficiency in learning complex spatial context, Chun and Jiang (1998) asked subjects to search for a T target among L distractors. Unknown to the subjects, some displays were occasionally repeated in the experiment. Such repetition led to a significant facilitation of search speed on repeated displays, even though subjects lacked explicit awareness of the repetition (Chun, 2000). Learning is observed only when the target location is fixed within a given repeated display. If the target location randomly changes from repetition to repetition, no learning is observed even though the global spatial layout remains the same (Chun & Jiang, 1998; Wolfe, Klempe, & Dahlen, 2000). This learning, known as "contextual cueing," is surprisingly powerful. It occurs after just five or six repetitions and lasts for at least a week (Chun & Jiang, 2003; Jiang, Song, & Rigas, *in press*).

What mechanism allows humans to search faster when a display is encountered for a second time? According to the instance theory (Logan, 1988), each visual search display leaves an implicit memory trace, an "instance." For

novel displays, subjects have to conduct standard, serial search to find the target. For a previously presented display, visual search becomes a race between standard search and memory retrieval. The latter occurs because the current display matches the memory instances laid down previously, so attention can be guided by past memory. Because instance-based attentional deployment is often faster than standard serial search, reaction time (RT) will be faster on repeated than on novel displays. The instance theory has been successful in accounting for visual procedural learning (Logan, 1988) and learning of repeated displays (Lassaline & Logan, 1993). It also provides a sound explanation for contextual cueing (Chun & Jiang, 1998).

A key component to contextual cueing is the retrieval of previous memory instances. In other words, the visual system must successfully match an incoming display with previous memory traces. The easier the match is established, the faster attention can be guided by previous memory. Yet how does the visual system connect the present display with previous memories? Must the current display match previous memory instances exactly? If the match does not need to be exact, to what degree can the differences be tolerated?

Previous studies show that differences in item identity can, under some conditions, be tolerated. For example, after they have searched from rotated 2s and 5s, subjects continue to search faster from the trained configuration that now contains distractors in a new shape (Chun & Jiang, 1998). Similarly, after subjects have learned a repeated spatial layout that contains black T and Ls, they continue to search faster among the repeated layout when the colors of all items have changed to white (Jiang & Song, *in press*). It appears that layout learning can be largely independent of the identity of distractors.

Differences in spatial layout pose a more serious problem to a successful match. After subjects have learned virtual three-dimensional (3D) displays viewed from a particular vantage point, learning fails to transfer when the same displays are viewed after a 30° to 90° rotation (Chua & Chun, 2003). Similarly, if only half of the items repeat their locations during training, the size of learning is much reduced (Chun & Jiang, 1998; Olson & Chun, 2002). Learning is preserved, however, if the entire display contracts or expands without deforming its global layout (Jiang & Wagner, 2004).

Taken together, these studies show that while an incoming display does not need to match a previously established memory trace exactly, maintaining good topographic matching is important for instance-based attentional guidance. However, no study has investigated the retrieval of memory instances when the incoming display deviates from the learned display. This study is designed to address this issue. We manipulated the degree of matching between a new display and a previously encountered display to study

whether instance retrieval can tolerate mismatch between the new input and the previous memory.

## Overview of experiments

In [Experiments 1](#) and [2](#), we studied memory *retrieval* on the basis of partial match between a new display and a well-learned display. We first trained subjects to learn a set of repeated visual search displays. These displays were repeated 28 times during training, allowing subjects to form a solid memory trace for each display. Following training, subjects were tested in a transfer phase that included new displays that matched the trained displays in 1 location, 2 locations, 3 locations, 4 locations, or all 12 locations. The “1 location” condition will be referred to as the *new* condition, in that except for the target location that matched the trained target location, all distractor locations were newly selected. The “12 location” condition will be referred to as the *old* condition, in that all items on the display, including the target and all distractors, repeated their locations from learning to transfer. The *new* and *old* conditions were thus the two baseline conditions, representing floor and ceiling performance, respectively. The other conditions – 2-, 3-, or 4-location matching – will be referred to as the *partial match* conditions and will be contrasted with the two baseline conditions.

It is important to note that the training session included only the *old* condition. In other words, all items (12 out of 12) on a display retained their locations during training. This served to establish a strong memory trace before the transfer phase started. This design allows us to examine the *retrieval* of already well-learned displays. In [Experiment 3](#) we modified the design to examine the *acquisition* of displays when only 3 out of 12 locations were preserved during learning.

## Experiment 1

In [Experiment 1](#), three conditions were tested during transfer: *old*, *new*, and *3-location old*. We carried out two versions of [Experiment 1](#): Both versions shared the same design sequence – with a training session followed by a transfer session – but differed slightly in the training procedure. [Experiment 1a](#) presented 12 colored items (1 T target and 11 L distractors) in four 3-item groups: yellow, green, blue, and red. Items repeated their colors as well as locations when a display was repeated. [Experiment 1b](#) presented 12 white items, so learning proceeded on the basis of spatial locations alone. The transfer phases of the two experiments were identical: 12 white items were presented on the search display such that color information was irrelevant during the transfer phase.

Both versions were tested because [Experiment 1a](#) can be considered as an intermediate step before [Experiment](#)

1b. **Experiment 1a** provided bottom-up cues about how the 12 items should be segregated into sets of three locations. Because the visual system is sensitive to color similarity (Driver & Baylis, 1989), the target and two other same-color distractors formed a single perceptual group. During transfer, these three items were repeated in the *3-location old* condition to potentially simplify the matching process. In **Experiment 1b**, bottom-up cues for grouping were absent during training, in which case the target may be grouped with any distractors with equal strength. Because there are many possible ways to divide 12 items into sets of 3 items, matching on the basis of partial overlap should be more difficult in **Experiment 1b** than **Experiment 1a**. **Figure 1** illustrates the design of the experiments.

**Method**

**Participants**

We recruited volunteers from around Harvard University. They were 18 to 35 years old and had normal color vision and normal or corrected-to-normal visual acuity. Fifteen subjects participated in **Experiment 1a** and 24 subjects participated in **Experiment 1b**.

**Equipment**

Participants were tested individually in a room with dim interior lighting. They viewed a computer screen from an unrestrained distance of about 57 cm, at which distance 1 cm corresponded to 1° visual angle.

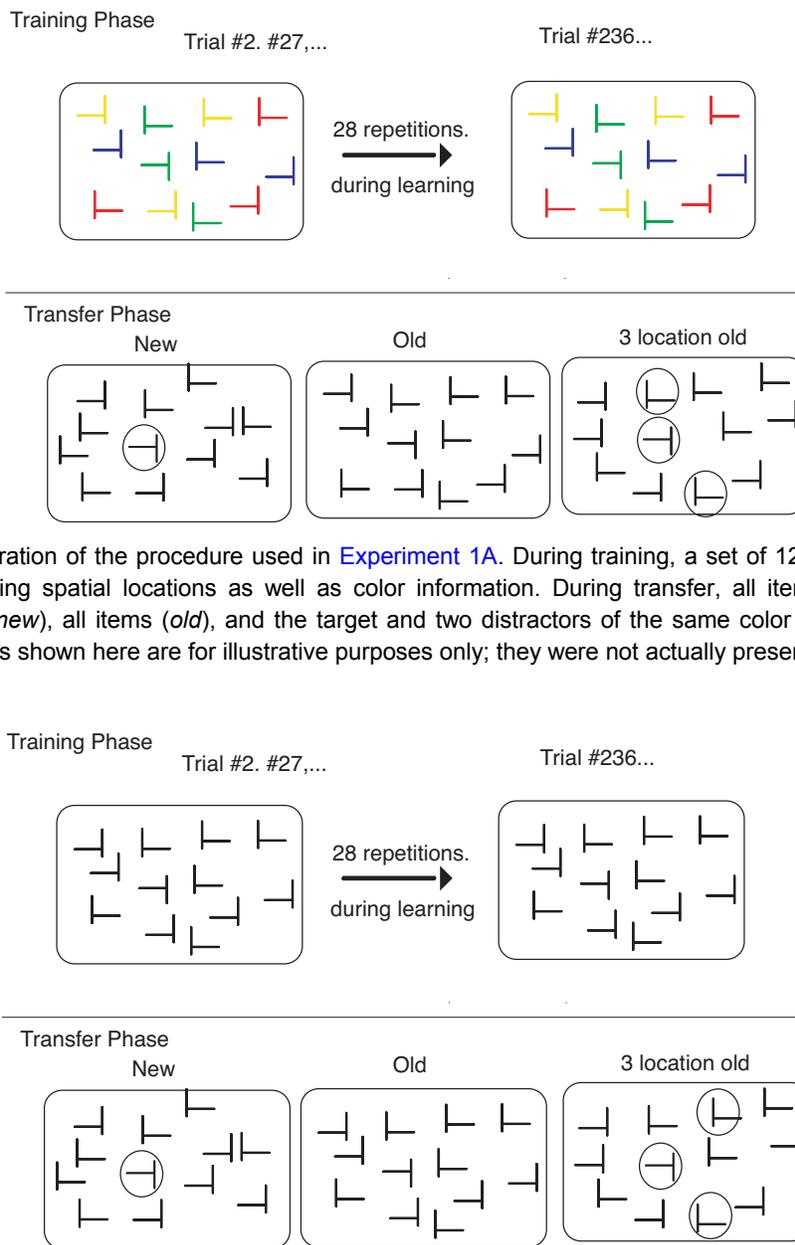


Figure 1A. A schematic illustration of the procedure used in **Experiment 1A**. During training, a set of 12 colored items was repeatedly presented 28 times, preserving spatial locations as well as color information. During transfer, all items were in white. The spatial locations of the target only (*new*), all items (*old*), and the target and two distractors of the same color during training (*3-location old*) were preserved. Dotted circles shown here are for illustrative purposes only; they were not actually presented.

Figure 1B. A schematic illustration of the procedure used in **Experiment 1b**. All items were presented in white throughout the experiment.

## Materials

Each visual search trial contained 12 items: 1 rotated T target and 11 rotated L distractors ( $0.7^\circ \times 0.7^\circ$ ), presented at randomly selected locations in a  $12 \times 8$  invisible grid matrix ( $23.4^\circ \times 15.6^\circ$ ). Subjects were instructed to search for the T target and press the left or right key to report its orientation. There was a small offset at the intersection of the Ls. The offset was  $0.2^\circ$  in [Experiment 1a](#); it was reduced to  $0.1^\circ$  in [Experiment 1b](#) because subjects complained that the Ls in [Experiment 1a](#) were too similar to the target T.

## Design

The experiment included two phases: training (28 blocks, 16 trials per block) and transfer (1 block, 48 trials). Prior to the first training block, 16 unique target locations were randomly chosen from the matrix. Each target was then presented with 11 randomly selected distractor locations to form a unique spatial layout. Each of the 16 spatial layouts was presented once per block and repeated 28 times. In [Experiment 1a](#), each display was divided into four color groups (red, green, yellow, and blue) of three items each. The colors for all groups were randomly chosen but preserved across blocks. In [Experiment 1b](#), all items were in white (see [Figure 1](#)).

The transfer phase immediately followed the training phase. It included 48 trials, randomly and evenly divided into three conditions: *old*, *new*, and *3-location old*. In both [Experiment 1a](#) and [Experiment 1b](#), all items were presented in white during the transfer phase. The *new* displays shared the target locations with the trained displays, but differed in their distractor locations. The *old* displays were the same as those seen during training, but color grouping was removed in [Experiment 1b](#). The *3-location old* displays shared three locations with the trained displays, while the other nine locations were randomly positioned. In [Experiment 1a](#), the repeated three locations included the target and two distractors that shared the target's color during training. In [Experiment 1b](#), the repeated three locations included the target and two randomly chosen distractors on the trained display.

The identity of the target (left or right T) was randomly determined on each trial such that a given repeated display was predictive of only where the target was, not what the target was.

## Trial sequence

Subjects pressed the spacebar to initiate each block. The search display was then presented until a response was made. Accuracy feedback ("Correct"/"Incorrect") was displayed immediately after each response. One second later the next trial commenced. Subjects were neither informed that displays would be repeating, nor were they given any special instructions before the transfer phase.

## Results

Although we did not test explicit recognition (which will be tested in [Experiment 2](#)), no subjects reported noticing the repetition of displays. We analyzed accuracy and RT. Trials with incorrect responses and trials with extreme RT falling outside of 3 SD of the mean of all trials for a given subject were excluded from the RT analysis. The latter criterion trimmed less than 2% of the complete dataset.

### 1. Experiment 1A: Training with color grouping

(1) *Training*. Mean accuracy ranged from 95% to 99% in different training blocks and was not significantly affected by block number,  $F < 1$ .

Mean RT was significantly affected by block number,  $F(27, 378) = 1.68, p < .02$ . RT became faster as training progressed ([Figure 2](#), left).

(2) *Transfer*. Accuracy in the transfer phase was above 95% and was not significantly different among the three transfer conditions,  $F(2, 28) = 1.02, p > .30$ .

Mean RT was significantly affected by transfer condition ([Figure 2](#), right),  $F(2, 28) = 7.46, p < .003$ . Planned contrasts showed that RT was significantly longer in the *new* than the *old* condition,  $t(14) = 2.68, p < .02$ , suggesting that subjects had learned the repeated displays during training. In addition, the *3-location old* condition was significantly faster than the *new* condition,  $t(14) = 2.87, p < .02$ , but not different from the *old* condition,  $t(14) = 0.62, p > .50$ . These results suggest that when perceptual grouping was provided during training, learning transferred completely to a display that repeated only 3 out of 12 locations.

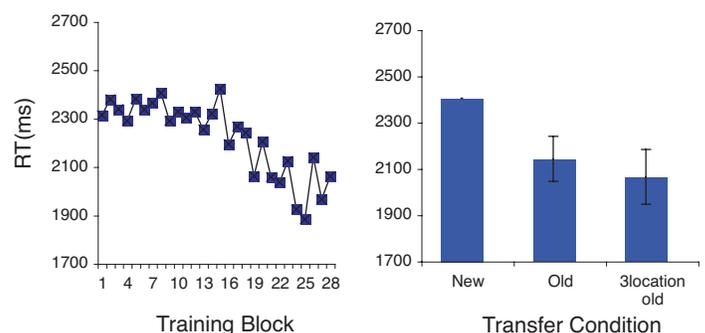


Figure 2. Results from [Experiment 1a](#). Left panel: training data. Right panel: transfer data. Error bars represent the standard error of the difference between each condition and the *new* condition.

### 2. Experiment 1b: Training without perceptual grouping

(1) *Training*. Mean accuracy ranged from 95% to 99% in different training blocks and was not significantly affected by block order,  $F < 1$ . Mean RT showed a significant improvement as the experiment progressed,  $F(27, 621) = 11.12, p < .001$  ([Figure 3](#), left).

(2) *Transfer*. Accuracy in the transfer phase remained high (above 95%) and was not significantly affected by condition,  $F(2, 46) < 1$ . Mean RT was significantly different among the three transfer conditions (Figure 3, right),  $F(2, 46) = 10.92$ ,  $p < .001$ . Planned contrasts showed that RT was significantly longer in the *new* than both the *old* condition,  $t(23) = 4.29$ ,  $p < .001$ , and the *3-location old* condition,  $t(23) = 2.44$ ,  $p < .03$ . The *old* and the *3-location old* conditions also differed significantly from each other, with the *old* condition faster,  $t(23) = 2.52$ ,  $p < .02$ . Thus, without color grouping during training, partial match on the basis of three repeated locations resulted in a significant, but incomplete, transfer of learning.

## Discussion

In Experiment 1, we first trained subjects on a set of repeated displays and then tested whether learning would transfer to displays that matched the trained displays in only 3 out of 12 locations. Compared with the *new* condition, the *3-location old* condition was more advantageous. This suggests that an exact match between a new display and the previous memory instance is not necessary for instance retrieval. How much benefit an incomplete match provided, however, depended in part on how strongly the matched locations were grouped together during training. The *3-location old* condition was as fast as the *old* condition in Experiment 1a, where the three repeated locations belonged to the same perceptual group during training. In Experiment 1b, where the three repeated locations were chosen completely at random from learned locations, the *3-location old* condition was slower than the *old* condition. This suggests that a stronger grouping cue modulates the degree of tolerance to mismatches. The interaction between Experiment 1a versus 1b and transfer condition (*old* vs. *3-location old*), however, was not significant,  $F(1, 37) = 2.41$ ,  $p > .13$ . Together, Experiments 1a and 1b suggest that first, an exact match between a new display and a previously learned display is not necessary for memory retrieval, and second, an exact match can be superior to an incomplete match at least sometimes. We will discuss the implications of these results in General discussion.

## Experiment 2

Experiment 2 extended Experiment 1b by testing additional partial match conditions. First, we would like to replicate the finding that a small number of overlap (e.g., 3 or 4 repeated locations out of 12) is sufficient to produce a transfer of learning. Second, we also wish to push the limit toward a lower number and estimate the minimal number of matching locations that still provides an advantage. To this end, we modified the transfer phase of Experiment 1b such that the new displays matched the trained displays in 1 location (*new*), 2 locations, 4 locations, or 12 locations (*old*). The training phase was identical to Experiment 1b.

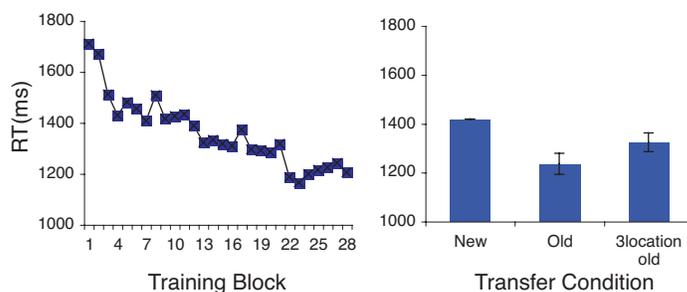


Figure 3. Results from Experiment 1b. Left panel: training data. Right panel: transfer data. Error bars represent the standard error of the difference between each condition and the *new* condition.

## Method

Twelve new subjects were tested in this experiment in a procedure similar to Experiment 1b. Subjects were first trained on 16 displays that repeated 28 times. Then during the transfer block, four conditions were tested. The *new* condition matched the trained displays only in the target's location (1 location match), and the *old* condition matched the trained displays in all 12 locations. The *2-location old* condition matched the trained displays in the target's location and one distractor's location. Finally, the *4-location old* condition matched the trained displays in the target's location and three randomly selected distractor locations. The transfer block contained 64 trials, randomly and evenly divided into the four conditions. Following the transfer block, we presented all 64 trials used in the transfer block again and asked subjects to determine whether they had seen any of the displays before. This last recognition block allowed us to assess whether learning in this experiment was explicit or implicit.

## Results

(1) *Recognition*. In the recognition phase of the experiment, the hit rate (reporting an old or partial-match condition as old) was .44, .41, and .41 for the *old*, *4-location old*, and *2-location old* conditions, respectively. These values were not significantly different from the false alarm rate (reporting the new displays as old) of .41, all  $ps > .20$ . Thus, any transfer we observed in this experiment was primarily a result of implicit learning.

(2) *Training*. Mean accuracy during training was high (95% to 99%) and was not significantly different in different blocks,  $F < 1$ . The training effect was shown primarily in RT (Figure 4, left). There was a significant main effect of block order on RT,  $F(27, 297) = 9.76$ ,  $p < .001$ .

(3) *Transfer*. Accuracy in the transfer phase ranged from 97% to 99%. It was not significantly affected by transfer conditions,  $F < 1$ .

Mean RT, however, was significantly affected by transfer condition (Figure 4, right),  $F(3, 33) = 14.62$ ,  $p < .001$ . In particular, the *new* condition was significantly slower than

the *old* condition,  $t(11) = 6.42$ ,  $p < .001$ , showing contextual cueing. Of the two partial match conditions, the *2-location old* condition was not significantly different from the *new* condition,  $t(11) = 0.30$ ,  $p > .70$ . It was significantly slower than the *old* condition,  $t(11) = 4.29$ ,  $p < .001$ , and slower than the *4-location old* condition,  $t(11) = 3.14$ ,  $p < .02$ . This suggests that repeating two locations was insufficient for any transfer to occur. Finally, the *4-location old* condition was significantly faster than the *new* condition,  $t(11) = 3.68$ ,  $p < .005$ , but significantly slower than the *old*,  $t(11) = 3.15$ ,  $p < .01$ . This suggests that repeating four locations resulted in a significant, but incomplete, transfer of learning.

## Discussion

Take together, [Experiment 1b](#) and [Experiment 2](#) showed that a minimum of about 3 matching locations (out of 12) was necessary for the retrieval of a previously learned memory instance. Why can retrieval operate on 3 or 4 matching locations but not 2 matching locations? A simple, perhaps oversimplified, calculation of display statistics helps us understand this observation.

Suppose we randomly sample 12 locations from a total of 96 locations (the parameters used in [Experiment 1b](#) and [Experiment 2](#)), and suppose we make two such random samplings. The likelihood that these two displays would, by chance, share at least  $N$  locations is

$$P(\text{overlap} \geq N) = 1 - P(\text{overlap} < N). \quad (1)$$

On this calculation, when the visual system detects two matching locations between an incoming display and a memory display, it has little basis to suspect that the two displays are the same: This could happen with nearly .5 probability for any two random displays. However, if the visual system detects a match in three locations, the likelihood that this happens by chance alone is reduced to .17. Increasing the match to four locations further reduces false alarm rate to .04. Thus, matching on the basis of three or four locations leads to a high probability of hits and a low probability of false alarms, whereas matching on the basis of two locations is much less accurate.

Thus, a simple calculation of display statistics provides a reasonably good account for why three or four matching locations but not two matching locations are sufficient for memory retrieval. It is unlikely, however, that the visual system relies exclusively on this simple statistical calculation. This is because this calculation predicts that matching would be about 96% accurate with four-location matching, but in actual data, four-location matching resulted in only a 56% transfer of learning. The discrepancy is understandable given that the simple statistical calculation makes assumptions about human visual perception that are unlikely true. In particular, it assumes that humans have perfect knowledge about the display characteristics (such as there are 96 total locations), and that humans can immediately detect the number of matching locations between two dis-

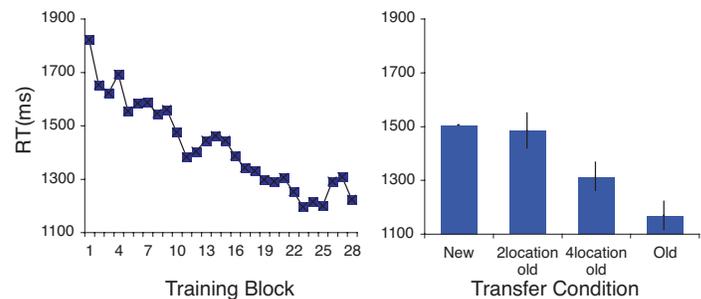


Figure 4. Results from [Experiment 2](#). Left: Training. Right: Transfer. Error bars represent the standard error of the difference between each condition and the *new* condition.

plays.<sup>1</sup> What the equation does provide though is a rationale for why two-location matching appears insufficient for successful retrieval of memory instances.

## Experiment 3

The first two experiments showed that once subjects had acquired a strong memory trace for repeated visual displays, learning partly transferred to displays that overlap with the trained ones in only three or four locations.

In this experiment, we investigated the effectiveness of partial match during *learning*. Specifically, we tested three conditions during the *training* phase: *old*, *new*, and *3-location old*. In the *old* condition, all items retained their locations when the display was occasionally repeated. Thus, the same exact display was repeatedly presented, once per block, for 28 times. In the *new* condition, the target location was repeated once per block, but all distractors changed their locations randomly. Finally, in the *3-location old* condition, three items (the target and two distractors) retained their locations when a display was repeated, while all other distractors were randomly positioned from block to block. [Figure 5](#) is a schematic illustration of displays.

Note that in the *3-location old* condition, subjects received no opportunity to establish a strong memory trace for the entire display. Instead, they must extract the three invariant locations among nine random locations from block to block. If the visual system relies on a more stringent criterion for the degree of matching during the initial training phase, then the presentation of nine randomly varying locations may be sufficient to disrupt or eliminate learning. Alternatively, if three-location repetition always satisfies the matching criterion, then subjects should learn from the *3-location old* condition.

## Method

### Participants

Fourteen subjects were tested in this experiment.

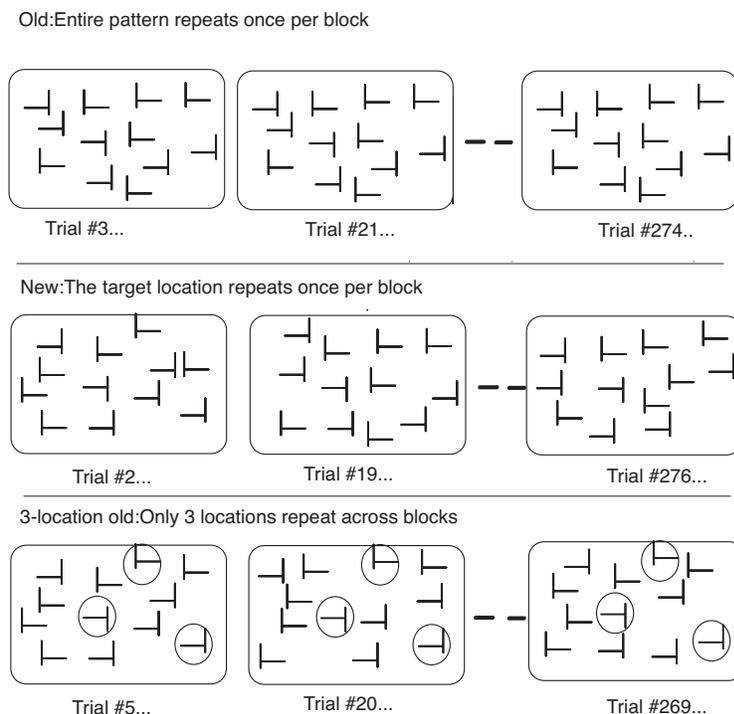


Figure 5. A schematic illustration of the three conditions tested during the training phase of [Experiment 3](#). Items are not drawn to scale; the dotted circles are for illustrative purposes only and were not shown on the actual experimental displays.

## Materials

The same materials as those used in [Experiment 1b](#) were used.

## Design

The experiment included only the training phase, which was divided into 28 blocks with 24 trials per block (8 trials per condition). Prior to the first block, 24 unique target locations were randomly chosen from a 12 x 8 invisible grid matrix. These locations were randomly and evenly assigned to three conditions: *new*, *old*, and *3-location old*. We then generated 11 random distractor locations for each target location and presented all 12 on the same search display. This resulted in 28 unique search displays per block. The target locations, but not the distractor locations, were repeated in the *new* condition across blocks. In the *old* condition, the entire display was repeated. Finally, in the *3-location old* condition, the target and 2 distractor locations were repeated across blocks while the other 9 distractor locations were randomly selected. The same 3 locations were shown once per block for 28 times.

Just as in [Experiment 1b](#), the identity of the target (left or right T) was randomly determined on each trial, so repeated distractor locations were predictive only of the target's location. We did not tell our subjects that some displays would be repeatedly presented. In post-experiment debriefing sessions, no subjects reported noticing the repeated displays.

## Results

Because each block contained only eight trials per condition, we binned four experimental blocks into one epoch to reduce noise in analysis. The entire experiment was thus divided into seven epochs.

Mean accuracy ranged from 95% to 97% in different epochs and was not significantly affected by training condition,  $F(2, 26) = 1.67$ ,  $p > .29$ , epoch,  $F(6, 78) = 1.36$ ,  $p > .20$ , or their interaction,  $F < 1$ .

[Figure 6](#) shows the group mean RT as a function of training condition and epoch. A repeated-measures ANOVA using condition (*old*, *new*, and *3-location old*) and

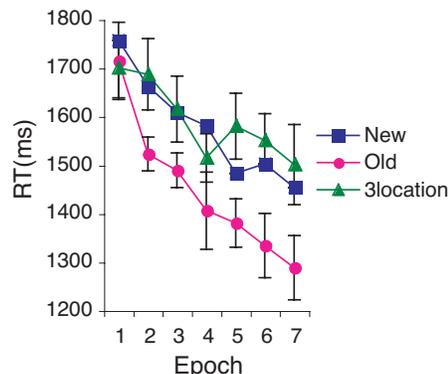


Figure 6. Results from [Experiment 3](#). Training data of three conditions: *new*, *old*, and *3-location old*. Error bars represent the standard error of the difference between each condition and the new condition.

epoch (1-7) as within-subject factors revealed a significant main effect of condition,  $F(2, 26) = 11.52$ ,  $p < .001$ , and a significant main effect of epoch,  $F(6, 78) = 5.72$ ,  $p < .001$ , but no interaction between the two,  $F(12, 156) = 1.43$ ,  $p > .10$ . Planned contrast showed that in Epoch 1, the three training conditions did not differ significantly from one another,  $F < 1$ . But in Epoch 7, they became significantly different,  $F(2, 26) = 5.22$ ,  $p < .02$ . In this epoch, RT was significantly faster in the *old* than both the *new*,  $t(13) = 2.86$ ,  $p < .02$ , and the *3-location old* condition,  $t(13) = 3.01$ ,  $p < .01$ . The *new* and the *3-location old* conditions were not significantly different from each other,  $t(13) = 0.61$ ,  $p > .50$ .

## Discussion

Is partial match on the basis of three repeated locations always sufficient for contextual cueing? The answer from [Experiment 3](#) is “no.” When subjects had to learn three repeated locations accompanied by nine randomly positioned locations, they failed to extract the invariant locations.<sup>2</sup> These results can be contrasted with those found in the first two experiments, where we observed a significant transfer of learning to new displays that matched the learned displays in three locations. Taken together, they suggest that to build up a stable memory representation, a stronger matching signal is required during the initial phase of learning. Once a strong memory trace for a repeated display is established, learning transfers even if a new display only minimally matches the previous memory.

## General discussion

Recent studies suggest that humans are severely impaired at representing visual details in conscious vision. For instance, only about three to four visual objects can be held in visual working memory. Yet at the same time, we are extremely efficient at extracting statistical regularities from visual displays, often in an implicit manner. Ever since Reber's pioneering studies on implicit learning (Reber, 1967, 1989), many studies have revealed a long list of invariant information that humans are sensitive to, including repeated spatial locations. Visual implicit learning may compensate for the severe limits in our conscious visual perception and working memory.

For such learning to occur, one must be able to match an incoming display with past memory. Yet on what basis does the visual system determine whether a match is found? Does a visual search display have to match exactly with previous memory for search to be guided by memory? Our study suggests that an exact match is unnecessary, at least late in the training phase. The degree of tolerance to non-matching information depends on whether a strong memory trace has already been established during the initial learning phase. The presentation of a small subset (e.g., 3 out of 12) of repeated locations is insufficient for learning.

But once a strong memory trace has been established, a new display that matches a learned display in only 3 or 4 (out of 12) locations can lead to a significant, albeit incomplete, transfer of learning. The tolerance also partly depends on how strongly the repeated locations were grouped initially during training: If the 3 repeated locations were perceived as one group during training, then repeating 3 locations can result in as much benefit as when repeating all locations. Given that subjects are unaware of display repetitions, it is extraordinary that successful instance retrieval can occur when a display matches a previous memory trace in only about 20-30% of locations.

What is the mechanism that allows the visual system to determine the match between an incoming display and a previous memory? As do other researchers (e.g., Lassaline & Logan, 1993), we believe that a similarity index is calculated: A new display is compared with previous memory instance. The more similar the two displays are, the more likely the visual system will rely on the retrieved memory to find the target. The calculation of similarity can be based on the entire configuration (how similar the whole display is to a previous memory configuration), or on a subset of the configuration, or even individual locations (Jiang & Wagner, 2004). Whether similarity is calculated on the basis of global display characteristics or on local features remains to be tested. Nonetheless, the degree of match needs to be higher during the initial learning phase before a strong memory trace is established.

## Conclusion

By training subjects on a set of repeated visual search displays and testing them on a partially matching new display, we found that humans can access a previous memory instance on the basis of about three or four matching locations. Search RT for partially matching displays is faster than that for new displays, although it is still slower than that for exactly repeated displays. Partial matching fails, however, when only three locations repeat during the initial training phase. We suggest that the visual system can tolerate mismatches between new displays and previous memory, especially late during training.

## Acknowledgment

This research was supported by National Institutes of Health Grant MH071788. JHS was supported by the Korea Foundation for Advanced Studies. We thank Sidney Burks for data collection, Patrick Cavanagh, Hing Yee Eng, Jeremy Wolfe, and an anonymous reviewer for comments.

Commercial relationships: none.

Correspondence author: Joo-Hyun Song.

Email: [jhsong@fas.harvard.edu](mailto:jhsong@fas.harvard.edu) or [yuhong@wjh.harvard.edu](mailto:yuhong@wjh.harvard.edu).

Address: 33 Kirkland Street, WJH 710, Cambridge, MA 02138.

## Footnotes

<sup>1</sup>We thank Jeremy Wolfe for raising these points.

<sup>2</sup>These results held even when we highlighted the three repeated locations with perceptual grouping cues. In a further experiment, we divided the 12 items into four groups of three, each group with a unique color. The three invariant locations on a given display were randomly assigned to a given color, such as red, and retained this color throughout the experiment. Even so, the *3-location old* condition was statistically indistinguishable from the *new* condition.

## References

- Biederman, I. (1987). Recognition-by-components: A theory of human image understanding. *Psychological Review*, *94*, 115-117. [PubMed]
- Chua, K. P., & Chun, M. M. (2003). Implicit scene learning is viewpoint dependent. *Perception and Psychophysics*, *65*, 72-80. [PubMed]
- Chun, M. M. (2000). Contextual cuing of visual attention. *Trends in Cognitive Science*, *4*, 170-178. [PubMed]
- Chun, M. M., & Jiang, Y. (1998). Contextual cueing: Implicit learning and memory of visual context guides spatial attention. *Cognitive Psychology*, *36*, 28-71. [PubMed]
- Chun, M. M., & Jiang, Y. (1999). Top-down attentional guidance based on implicit learning of visual covariation. *Psychological Science*, *10*, 360-365.
- Chun, M. M., & Jiang, Y. (2003). Implicit, long-term spatial contextual memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *29*, 224-234. [PubMed]
- Driver, J., & Baylis, G. C. (1989). Movement and visual attention: The spotlight metaphor breaks. *Journal of Experimental Psychology: Human Perception and Performance*, *15*, 448-56. [PubMed]
- Fiser, J., & Aslin, R. N. (2001). Unsupervised statistical learning of higher-order spatial structures from visual scenes. *Psychological Science*, *12*, 499-504. [PubMed]
- Jiang, Y., & Song, J. -H. (in press). Hyper-specificity in visual implicit learning: Learning of spatial layout is contingent on item identity. *Journal of Experimental Psychology: Human Perception and Performance*.
- Jiang, Y., Song, J. -H., & Rigas, A. (in press). High-capacity spatial contextual memory. *Psychonomic Bulletin and Review*.
- Jiang, Y., & Wagner, L. C. (2004). What is learned in spatial contextual cueing: Configuration or individual locations? *Perception and Psychophysics*, *66*, 454-463. [PubMed]
- Lassaline, M. E., & Logan, G. D. (1993). Memory-based automaticity in the discrimination of visual numerosity. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *19*, 561-581. [PubMed]
- Li, F. F., VanRullen, R., Koch, C., & Perona, P. (2002). Rapid natural scene categorization in the near absence of attention. *Proceedings of the National Academy of Sciences U.S.A.*, *99*, 9596-9601. [PubMed][Article]
- Logan, G. D. (1988). Toward an instance theory of automatization. *Psychological Review*, *95*, 492-527.
- Miller, J. (1988). Components of the location probability effect in visual search tasks. *Journal of Experimental Psychology: Human Perception and Performance*, *14*, 453-471. [PubMed]
- Nissen, M. J., & Bullemer, P. (1987). Attentional requirements of learning: Evidence from performance measures. *Cognitive Psychology*, *19*, 1-32.
- Olson, I. R., & Chun, M. M. (2001). Temporal contextual cueing of visual attention. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *27*, 1299-1313. [PubMed]
- Olson, I. R., & Chun, M. M. (2002). Perceptual constraints on implicit learning of spatial context. *Visual Cognition*, *9*, 273-302. [PubMed]
- Palmer, S. E. (1999). *Vision science: Photons to phenomenology*. Cambridge, MA: MIT Press.
- Reber, A. S. (1967). Implicit learning of artificial grammars. *Journal of Verbal Learning and Verbal Behavior*, *5*, 855-863.
- Reber, A. S. (1989). Implicit learning and tacit knowledge. *Journal of Experimental Psychology: General*, *118*, 219-235.
- Thorpe, S., Fize, D., & Marlot, C. (1996). Speed of processing in the human visual system. *Nature*, *381*, 520-522. [PubMed]
- Wolfe, J. M., Klempen, N., & Dahlen, K. (2000). Postattentive vision. *Journal of Experimental Psychology: Human Perception and Performance*, *26*, 693-716. [PubMed]