Visual working memory for briefly presented scenes

Kristine Liu
Department of Psychology, Wellesley College, Wellesley, MA, USA

Yuhong Jiang
Department of Psychology, Harvard University, Cambridge, MA, USA

Previous studies have painted a conflicting picture on the amount of visual information humans can extract from viewing a natural scene briefly. Although some studies suggest that a single glimpse is sufficient to put about five visual objects in memory, others find that not much is retained in visual memory even after prolonged viewing. Here we tested subjects’ visual working memory (VWM) for a briefly viewed scene image. A sample scene was presented for 250 ms and masked, followed 1000 ms later by a comparison display. We found that subjects remembered fewer than one sample object. Increasing the viewing duration to about 15 s significantly enhanced performance, with approximately five visual objects remembered. We suggest that adequate encoding of a scene into VWM requires a long duration, and that visual details can accumulate in memory provided that the viewing duration is sufficiently long.

Keywords: visual working memory, natural scenes, capacity limitation, change detection

Introduction

Ten objects seem trivial next to the scores, even hundreds, of objects that we encounter in any given glance of our surroundings. Close your eyes, turn your head in any direction, open and close them as fast as possible. How much do you remember? You probably remember the gist of the scene, but do you remember all the objects? In this study we suggest that a single glimpse is too brief to get even one object in visual working memory (VWM) but prolonged viewing allows memory to accumulate. As intuitive as these messages are, both claims are inconsistent with many influential studies conducted in the past decade.

A number of studies suggest that we get the gist of a scene rapidly (Bar, 2004; Intraub, 1981; Potter, 1976; Potter, Staub, Rado, & O’Connor, 2002). Potter and colleagues show that even when a series of scenes are presented at a rate of 113 ms per item, people are still quite good at detecting a presupposed scene. In addition, visual details of a scene, such as whether a scene contains animals or not, can be extracted within about 150 ms (Thorpe, Fize, & Marlot, 1996; Li, VanRullen, Koch, & Perona, 2002). The details of a scene are not only encoded for immediate visual categorization but also retained in memory for later recall and recognition (VanRullen & Koch, 2003). For example, VanRullen and Koch (2003) presented subjects with common scenes such as bedroom or farm scenes generated by computers. Each scene contained 10 objects and was presented for 250 ms and masked. They found that subjects were able to freely recall the names of about 2.5 objects. Furthermore, when forced to choose from an array of objects, subjects were able to pick out additional 2.5 objects. Even the other five objects not recalled or recognized stayed in visual memory, resulting in negative priming in a later picture–word matching task. Thus, a single glimpse at a visual scene puts about five visual objects in working memory and five others in implicit memory.

Not all studies are so optimistic about the amount of visual details that the visual system extracts. When asked to detect a visual change between two cyclical scenes that differ by one change, people often spend several seconds inspecting the images, even when the change is conspicuous once pointed out (Rensink, O’Regan, & Clark, 1997; Scholl, Simons, & Levin, 2004). Our inability to detect visual changes, known as change blindness, suggests that visual details of a scene are not well retained in memory, even when the image is viewed for several seconds (Most et al., 2001; Simons & Chabris, 1999; Wolfe, Klempen, & Dahlen, 2000). O’Regan (1992) proposed that the visual system does not retain detailed visual information in working memory. When we need to use information from the outside world, we do not consult our immediate memory; we simply open our eyes and look.

In this study, we wish to measure the amount of visual information that stays in VWM after people view a briefly presented scene. This is a question to which neither lines of research reviewed above have provided conclusive answers. On one hand, studies that flash a scene and test subjects’ recall and recognition often permit the usage of nonvisual memory stores. In VanRullen and Koch’s (2003) study, for example, subjects were asked to verbally name objects during free recall and to match the name of an object with its picture during recognition, thereby recruiting verbal working memory. This may have led to the conclusion that humans have a larger and more efficient VWM capacity than they actually have. On the other hand, change blindness studies might have underestimated the amount of visual details that humans can extract from
viewing a scene (Hollingworth, 2004). A failure to detect a change does not necessarily mean that people did not retain the sample scene in memory. Rather, subjects might have failed to compare the sample and test scenes (Angelone, Levin, & Simons, 2003; Most, School, Clifford, & Simons, 2005), or their memory for the sample scene might have suffered interference by the test scene (Landman, Spekreijse, & Lamme, 2003).

We conducted two experiments. In Experiment 1, we presented subjects with a sample scene briefly (250 ms and masked). After a short retention interval (1000 ms), a test screen was presented. The test screen might contain 20 choice objects for subjects to pick 10 that matched the sample. In Experiment 2, the test screen contained four choice objects for subjects to pick one that matched the sample. In an attempt to reduce the support of verbal working memory, verbal encoding was suppressed either through articulatory suppression (Baddeley, 1986), or by testing people on filler objects that shared the same name as the target objects. We also reduced gist-based recognition by including on the test display only objects that fit with the sample scene.

To narrow down factors that limited performance, in Experiment 2 we tested control subjects who viewed a sample scene for as long as they wanted. If these subjects also performed as poorly as subjects in the brief-viewing condition, then the limitation must have originated from an inability to retain scene details in VWM. In contrast, if performance increased as a function of viewing duration, then the limitation we saw in the brief-viewing condition must have originated from inadequate perception of the scene.

**Experiment 1**

**Method**

**Participants**

Twelve naïve subjects (18–26 years old) from Harvard University participated for pay.

**Scene stimuli**

We selected 10 photographed scenes, including a kitchen, a bathroom sink, an empty bedroom, an empty living room, a TV room, a city street, a farm, an office desk, a coffee table, and the inside of a refrigerator. We also selected 20 plausible objects for each scene. The 20 objects were randomly divided into two sets of 10 objects each. Each set was then placed at plausible locations in its associated scene, producing two sets of 10 scenes each. Half of the subjects viewed one set of the scenes and the other subjects viewed the other set of the scenes. Figure 1 shows a pair of two scenes, tested on separate subjects.

**Choice stimuli**

Following the presentation of the sample scene, a test display containing all 20 plausible objects for that scene was presented. The objects were presented at randomly selected locations from a 5 × 4 matrix (25° × 20°). Ten objects matched the sample scene and are the “targets,” whereas the other 10 objects were the
“fillers.” The filler objects were the ones taken from the alternative version of the same scene, which resulted in objects that fit the gist of the scene. The same objects were targets for half of the subjects and fillers for others, and vice versa across all subjects. This design allowed us to control for guessing biases. For example, suppose that “pillow” is more highly associated with a bedroom scene than “trashcan” is, and suppose subjects would make guesses based on association strength, then that guess (“pillow”) would be correct for one subject but wrong for another.

**Trial sequence**

On each trial a sample scene was presented for 250 ms and masked by a grid of color patches for 250 ms. Then after a blank interval of 1000 ms, the test screen containing 20 choice objects was presented. Subjects pressed the letter at the position of an object that matched their memory; that object was then marked out with a bright red circle. Subjects then rated their confidence for that object on a 1–7 scale, with 1 being *not at all confident* and 7 being *very confident*. This choice response continued until 10 objects were marked out and rated, at which point the choice screen was erased. The next trial started after a key press.

**Conditions**

To discourage verbal labeling of the objects, half of the observers were tested using articulatory suppression in which they repeated out aloud “teddy bear” throughout the experiment (Baddeley, 1986). The remaining subjects were silent during the experiment.

**Results**

We analyzed recognition accuracy and confidence rating separately for different subject groups (articulatory suppression or silent). Figure 2 shows the average for the two groups.

Using subject group as a between-subject factor and recognition order as a within-subject factor, an ANOVA showed a significant main effect of recognition order on accuracy, $F(9,90) = 3.71$, $p < .001$, and on confidence rating, $F(9,90) = 54.95$, $p < .001$. Objects selected earlier in the recognition phase were associated with higher accuracy and confidence. The main effect of articulatory suppression was not significant on accuracy, $F(1,10) = 1.70$, $p > .20$, or on confidence rating, $F < 1$, ns, and the interaction effect was not significant in either measure, $F s < 1$, ns.

Because articulatory suppression had a negligible effect on performance, in the following analysis we pooled data across all subjects. A one-sample $t$-test showed that in accuracy the first object that subjects chose during recognition was quite accurate ($M = 75\%$, $SE = 6.4\%$), significantly above what would be expected based on chance ($50\%$), $t(11) = 3.85$, $p < .003$. However, objects chosen at later positions (from 2 to 10) were quite poor, with an average of $48.4\%$ for those nine positions. Given that subjects had selected the first object at $75\%$ correct, the chance for the second choice would be $[75\% / 9] + [25\% / 10] = 48.7\%$. This value was not significantly different from the average of choices 2–10 ($p > .50$), and was not significantly lower than each of the 9 choices, all $p$ values $>.30$. Thus, subjects’ first choice was made with good accuracy, but the later choices appeared to be random guesses.

**Discussion**

The results of Experiment 1 suggest that after briefly viewing a complex scene, subjects recognized at most one object accurately. These findings are apparently inconsistent with a previous study that used a similar procedure but allowed subjects to recall the names of exposed objects (VanRullen & Koch, 2003). There were several differences between these studies that individually or jointly may account for the discrepancy in results.
First, the two studies differed in the recognition procedure. Although both studies used 20 objects on the choice screen, half new and half old, the procedure for controlling for prior association biases was different. In VanRullen and Koch’s (2003) study, the new objects were selected “so that they could have normally been present in the context of the scene” (p. 76). Although this procedure should make both new and old objects plausible within a scene, there could still be differences in the degree of plausibility. If the old objects were more highly associated with the scene than the new objects were, then subjects could make educated guesses about the target objects. In our study, new objects for one subject were target objects for another, so educated guesses would lead to a correct response for one subject and an incorrect response for another. This difference in recognition procedure would lead to lower recognition accuracy in our study than in theirs.

Second, the two studies differed in the degree to which nonvisual cues could contribute to later recognition. VanRullen and Koch’s (2003) study allowed verbal memory to enhance performance. This was ensured in two ways. First, the recognition memory task was conducted after a verbal recall task, a procedure that would encourage subjects to rely on verbal memory. Second, the new and old objects were selected such that they would have distinct names. In our study, in contrast, recognition came immediately after each scene, which might have discouraged subjects from using verbal memory. Furthermore, in our study the new objects tended to share names with the old ones. Reducing the contribution of verbal memory would lead to lower recognition accuracy in our study than in VanRullen and Koch’s study.

Finally, the stimuli used in these two studies were different. Both studies used computer software to place individual objects against a photographed background. However, certain visual cues—such as shadows—were added on the objects in VanRullen and Koch’s (2003) study to make the scene look more natural. Our study, however, did not preserve subtle cues like this, although cues such as occlusion were included. Differences in exactly how stimuli were created could make one set of stimuli more realistic than another. It is possible that recognition memory would be higher in more realistic images, again resulting in lower performance in our study than in VanRullen and Koch’s.

Perhaps the differences listed above can jointly account for the discrepancy between our study and VanRullen and Koch’s (2003). Each study has its own merit. In everyday vision, it is likely that subjects do have access to verbal memory and are allowed to make educated guesses about what objects appear in a scene. In this sense, the presence of these cues in VanRullen and Koch’s study would be a good approximation to subjects’ performance. However, because we are interested in the amount of information held in visual memory per se, it is important that verbal cues and educated guesses be reduced or eliminated. In this sense, our study gives a good estimation about how much is retained in VWM from a briefly glanced scene.

### Experiment 2

A surprising result from Experiment 1 is that a single glimpse at a scene allowed at most one object to be retained in VWM. Experiment 2 is designed to address a few remaining questions from Experiment 1. First, it is unclear whether the poor performance resulted from a limitation in the encoding phase or in the retention phase. That is, accuracy could be low either because the presentation duration was too brief for adequate encoding of scene details, or that subjects encoded all objects but were unable to retain them in working memory. To address this question, we tested two groups of subjects in Experiment 2. One group viewed each sample display for 250 ms, as in Experiment 1. The other group was allowed to view each sample display for as long as they wanted before the display was masked. A long presentation duration ensured that subjects had successfully encoded a scene, so the remaining limitation would correspond to a retention limitation (Eng, Chen, & Jiang, in press).

A second change we made was the recognition procedure. By requiring subjects to pick out all 10 sample objects sequentially, Experiment 1 might have underestimated the amount of information subjects extracted. This procedure was analogous to Sperling’s (1960) “whole report” procedure, in which memory for a display decayed during the lengthy recognition phase. It is possible that subjects might have retained more visual objects, but as they selected the first object from the sample, their memory for the other objects had decayed. To make a more accurate estimation of subjects’ memory accuracy, in Experiment 2 we used a partial report procedure during recognition. In this procedure, a randomly selected sample object was presented on the choice screen along with three other fillers. Subjects were asked to select the sample object. Because subjects only needed to make a single recognition rather than a series of 10 recognitions, this procedure minimized memory decay during recognition.

Third, to further minimize the effect of verbal memory, on the choice display we presented a filler that matched the target object’s name. For example, suppose there was a wooden chair on the sample scene, then the choice display would contain two wooden chairs of different designs and two desks of different designs. If subjects had no memory for the chair, then they would randomly choose from the four choices. If subjects had verbal memory but not visual memory for the chair, then they
would reject the two tables but randomly choose from the two chairs. Finally, if subjects had visual memory of the chair, then they would correctly identify the target. Thus, by measuring accuracy and the type of errors that subjects committed, we could estimate the relative contribution of visual and verbal working memory on recognition performance.

Finally, we wanted to examine how the presence of a congruent scene background affected VWM for individual objects in the scene. Because a congruent scene facilitates the perception of objects (Bar, 2004; Biederman, Mezzanotte, & Robinowitz, 1982; Davenport & Potter, 2004), it might enhance VWM for these objects. Alternatively, a display without its background layout is visually simpler. This might allow individual objects to be more easily segregated from the background, resulting in better memory.

In short, by manipulating the duration that subjects could view the sample display, this experiment separates encoding limitations from retention limitations. The partial report procedure allowed us to get a more accurate estimation of memory and to identify the relative contribution of visual and verbal working memory. Finally, by testing subjects on displays with or without the background setting, we can assess the role of background scene on VWM for individual objects.

Method

Participants

Thirty-two new observers participated in this experiment for payment. They were assigned randomly and evenly into four groups, as specified below.

Stimuli

The same 20 scenes created for Experiment 1 were used for half of the subjects (“scene context present”), where subjects viewed objects with their background setting. For the other half of the subjects, all 20 scenes were altered by removing the background (“scene background absent”). That is, the context (a kitchen counter, a desk, the fridge, etc.) was removed entirely save for the objects. The 10 objects in each scene were thus situated on a white background in the same positions they previously occupied in the scene context-present condition. To increase the number of trials per subject, we tested each subject on all 20 scenes (two sets of 10 different gist) in a randomly determined order.

Procedure

In each of the scene context condition, half of the subjects viewed each scene for 250 ms and the other half viewed it for as long as they wished. In the latter case, the sample scene would stay on the screen until the subjects hit a key. After the stimulus presentation, a color mask (see Experiment 1) was flashed for 250 ms. A retention interval of 1000 ms later, a test display including four choice objects was presented. The four objects included a target object that matched one of the sample objects. The other three were fillers. These fillers were constructed as follows. One filler, the “cousin” of the target, shared the same name as the target but differed in visual details. The other two fillers, the “new” objects, had a different name than the target object, although the two fillers themselves were of the same name. They were not actually presented on the preceding display, but they were plausible objects for that scene. In fact, one of the new objects was presented in the alternate version of the same gist. For instance, in the example of the city street scene, if the target was the traffic light, the four choices offered would be the target, another traffic light, a street lamp that belonged to the alternative city street scene, and a completely novel street lamp. The four objects were presented at randomly selected locations from a $2 \times 2$ invisible matrix ($6^\circ \times 6^\circ$) centered on fixation. Subjects were told to press the digit next to the target object that matched their memory. Figure 3 shows a sample test trial.

Results

We categorized each response as a hit or a false alarm, and the false alarm trials were further divided into “cousin” and “new,” depending on whether subjects erred on the object that shared the target’s name or the new objects. Figure 4 shows the results, separately for each presentation duration and scene context condition.

We first conducted an ANOVA on the overall accuracy (hit rate), with scene context (present vs. absent) and presentation duration (250 ms or self-paced) as between-subject factors. We observed a significant main effect of scene context, with higher accuracy in the context-absent condition, $F(1,28) = 5.88, p < .022$. The main effect of presentation duration was also significant, with higher accuracy in the self-paced condition, $F(1,28) = 48.13, p < .001$. The interaction between the two factors was not significant, $F < 1$, suggesting that increasing the presentation duration was beneficial for both context-present and context-absent conditions. On average, subjects in the self-paced condition took 13.7 s to view an image when the scene context was present and 13.2 s to view an image when the scene context was absent. These two values were not significantly different from each other, $t(14) = 0.21$, $p > .50$. Of the 16 individuals who were tested in the self-paced condition, we observed a significant correlation between viewing duration and accuracy, $r = .54, p < .029$. That is, those observers who viewed the sample scene longer also had higher accuracy.

How much did verbal memory contribute to performance in this experiment? If subjects only remembered the objects’ names, they would be more likely to commit false alarms on “cousin” objects that shared the target’s name than on “new” objects. An ANOVA on
error type (cousin vs. new), scene context (present vs. absent), and presentation duration (250 ms or self-paced) revealed no main effect of error type, $F(1,28) < 1, ns$. In addition, error type did not significantly interact with other experimental factors, all $F$ values $< 1.1$, $p$ values $> .30$. Thus, there was no evidence that subjects succeeded in the task simply by remembering the names of the viewed objects.

Finally, we estimated the number of visual objects held in VWM in each condition. Suppose the number of items retained was $N$, where $N$ is less than the total number of items presented (10). If the target object happened to be among the $N$ items retained, then accuracy would approach 100%. But if the target object happened to be the other items not retained in memory, then accuracy would be at chance (25%). Thus, the following equation holds:

$$\text{Accuracy} = \left[ N \times 100\% + (10 - N) \times 25\% \right] / 10$$

Plugging in a subject’s accuracy allowed us to solve $N$ for that subject. When a consistent context was presented, subjects retained about 0.67 objects when viewing a display for 250 ms and 5.33 objects when viewing a display for a longer duration. When a scene context was absent, subjects retained about 2.1 objects when viewing a display for 250 ms and 7.41 objects when viewing a display for a longer duration.

![Figure 3](https://jov.arvojournals.org/pdfaccess.ashx?url=/data/journals/jov/933511/)

Figure 3. Sample displays and trial sequence used in Experiment 2.

![Figure 4](https://jov.arvojournals.org/pdfaccess.ashx?url=/data/journals/jov/933511/)

Figure 4. Results from Experiment 2. FA: False alarm. The error bars show standard error of the mean across subjects.
Discussion

Experiment 2 showed that when objects were presented against a plausible scene for 250 ms, subjects retained about 0.7 objects in visual memory. This estimate was similar to that obtained in Experiment 1, suggesting that even with a partial report procedure, subjects still revealed very poor memory. However, recognition performance improved significantly in the self-paced condition. When allowed to view a scene for as long as they wanted (on average, for 13 s), subjects recognized about 5.3 objects out of 10. This difference suggests that the poor performance observed in the short viewing condition was primarily due to encoding failure rather than retention failure. When a scene was viewed for 250 ms, most visual details were not encoded into VWM. Thus, although 250 ms may be sufficient for people to perceive semantic gist of a scene (Potter, 1976), or to judge whether a scene contains animals (Li et al., 2002; Thorpe et al., 1996), it is not long enough for most visual details to enter VWM.

Increasing the encoding duration significantly enhanced the number of visual objects retained. This finding suggests that details of a visual scene can accumulate in visual memory, a conclusion inconsistent with that derived from change blindness studies. It is, however, consistent with an alternative view of visual memory held by Hollingworth, Henderson, and their colleagues (for a review, see Hollingworth, in press). In one study, Hollingworth (2004) allowed subjects to view a computer-generated scene for 20 s. Subjects were then presented with a scene that was slightly altered. Hollingworth found that subjects were able to detect changes from the original scene to the altered version, even when the changes involved visual details such as the rotation of a single object. Hollingworth and Henderson (2003) further showed that subjects were still able to detect changes when the lag between the initial viewing and later testing was as long as a day. This suggests that prolonged viewing of a display allowed visual details to be retained in visual memory, supporting immediate or long-term recognition. Our results are consistent with Hollingworth’s findings in showing that accumulation of visual details in memory was possible. In addition, we showed that such accumulation took substantially longer than a single glimpse.

What kind of memory allowed subjects to retain about 5.3 objects out of 10? Recent studies on VWM suggest that it has a capacity of about four for simple features (e.g., colors) or about two for complex features (e.g., random polygons; Alvarez & Cavanagh, 2004; Luck & Vogel, 1997). Few studies have estimated the capacity of VWM for everyday objects like those used in our study. It is possible that all of the 5.3 objects were held in VWM, as this number is close to the theoretical upper limit of VWM (Cowan, 2001). However, it is also possible that long-term visual memory was recruited. Our study cannot unambiguously separate the contribution of VWM from visual long-term memory. Because real objects have access to long-term memory, it will be difficult to eliminate contributions from visual long-term memory. Future studies that use novel objects may better address this issue.

Finally, results from Experiment 2 showed that compared with no background context, placing an array of 10 objects on a congruent scene impaired the number of objects retained, both for brief viewing and for prolonged viewing. This finding may seem to be at odds with previous findings that observed an enhancement from congruent context (e.g., Biederman et al., 1982). However, in studies examining context effect, congruent context was often compared with incongruent context rather than with no context. When compared with isolated objects without any context, a congruent scene increased visual complexity of the display and the difficulty to parse individual objects. Consistent with this interpretation, Davenport and Potter (2004) found that performance in a consistent-context condition was worse than a no-context condition.

General discussion

Previous studies suggest that a single glimpse is enough to know the semantic gist of a scene (Potter, 1976). In addition, even visual details, such as whether a scene contains animals or not, can be extracted within about 150 ms (Li et al., 2002; Thorpe et al., 1996). Further, when probed in a later recall and recognition task, subjects could explicitly retain approximately five to six visual objects after viewing a masked scene for 250 ms (VanRullen & Koch, 2003). Such stunning efficiency stands in contrast to surprising limitations seen under other testing conditions. Large visual changes can often go unnoticed, even in natural viewing conditions or social interactions (Levin & Simons, 1997; Simons & Chabris, 1999; Simons & Levin, 1998). These findings have led to contrasting views about the efficiency of the visual system at extracting and accumulating visual details from a scene. The present study shows that there are elements of truth to both views, although neither is complete.

First, our finding suggests that when verbal memory and educated guesses were reduced, the visual system was able to extract very few details from a single glimpse at a scene. Recognition performance was poor, whether objects were presented without a context (Experiment 2) or placed against a consistent context (Experiments 1 and 2). In a follow-up study, we used real scene photographs and presented these for 250 ms or 16 s. After a brief retention interval, subjects were shown an altered version of the photograph and must discriminate a region that had changed from a region that had not changed. Performance
was at chance with the 250-ms viewing duration. Thus, even with real photographs of natural scene images, a single glimpse does not supply a lot of information to visual memory. This is not to deny that a single glimpse is enough for getting the semantic gist, neither does our result negate the possibility that enough visual statistics might have been extracted to make general categorization (such as the presence of animal vs. vehicle). In fact, in everyday vision, people can rely on nonvisual information (e.g., verbal labels) and make educated guesses about which objects are likely present in a scene. This means that under those conditions people may perform better than what is estimated in this study (VanRullen & Koch, 2003). However, such representation is unlikely held entirely in VWM, and it lacks visual precision.

Prolonged viewing of a scene changes the results significantly. When subjects were allowed to control their viewing duration of a scene, they often spent about tens of seconds scrutinizing the scene, resulting in substantially better visual memory. Subjects could then retain about five objects buried in a scene or seven objects without a scene. Such performance was excellent considering that VWM could hold only about four simple objects or two complex objects. Contrary to the view that nothing accumulates in visual memory (Rensink et al., 1997; Simons, 1996; but see Simons & Rensink, 2005), our results suggest that the visual system is capable of retaining detailed visual information (see also Hollingworth, 2004). In the same follow-up study using natural scenes as mentioned earlier, subjects were able to pick out a changed region with 63% accuracy (chance was 50%) after viewing the scene for 16 s. Thus, provided there was enough time to view a visual scene, details of a scene do accumulate in visual memory.

Our study, together with previous studies, suggests that the visual system has two routes to make sense of the complex world: a fast route to extract semantic gist (Potter, 1976), high-level categorization (Grill-Spector & Malach, 2004; Li et al., 2002), or perceptual priming (Schacter & Buckner, 1998) and a slow route to accumulate veridical visual details (Hollingworth, 2004; Rensink et al., 1997). These findings are consistent with the “reverse hierarchy theory” postulated by Hochstein and Ahissar (Hochstein & Ahissar, 2002; Ahissar & Hochtein, 2004). Because natural scenes often stay remarkably stable, people often have adequate time to extract visual details if they wish to do so. The two routes thus operate in parallel in everyday vision, allowing us to navigate visual environment with ease.

**Acknowledgments**

This research was supported by the National Science Foundation 0345525 (YJ). We thank Rufin VanRullen for help with stimuli used in Experiment 1.

**Commercial relationships:** none.
**Corresponding author:** Yuhong Jiang.
**Email:** yuhong@wjh.harvard.edu.
**Address:** 33 Kirkland Street, WJH 820, Cambridge, MA 02138.

**References**


