Published in final edited form as:
J Exp Psychol Hum Percept Perform. 2009 February ; 35(1): 28–38. doi:10.1037/a0013624.

Detecting and Remembering Simultaneous Pictures in a Rapid Serial Visual Presentation

Mary C. Potter and **Laura F. Fox**
Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology

Abstract

Viewers can easily spot a target picture in a rapid serial visual presentation (RSVP), but can they do so if more than 1 picture is presented simultaneously? Up to 4 pictures were presented on each RSVP frame, for 240 to 720 ms/frame. In a detection task, the target was verbally specified before each trial (e.g., *man with violin*); in a memory task, recognition was tested after each sequence. Target detection was much better than recognition memory, but in both tasks the more pictures on the frame, the lower the performance. When the presentation duration was set at 160 ms with a variable interframe interval such that the total times were the same as in the initial experiments, the results were similar. The results suggest that visual processing occurs in 2 stages: fast, global processing of all pictures in Stage 1 (usually sufficient for detection) and slower, serial processing in Stage 2 (usually necessary for subsequent memory).

Keywords

picture perception; picture memory; target detection; RSVP; search

As people look around their normal environment, they take in the scene in a series of fixations lasting about 250 ms. Just how much information can be extracted from each fixation, and how well can it be remembered later? Recent studies have suggested not only that a scene can be understood within such a glimpse, but also that a target can be detected among as many as four simultaneous scenes presented briefly, at little or no additional cost (Rousselet, Thorpe, & Fabre-Thorpe, 2004b). In the present study we investigate this claim using two tasks, detection and later memory.

The ability to detect a target almost as well among several items as when only one item is presented suggests some capacity for processing multiple items in parallel. Indeed, studies of the monkey visual system using single-cell recordings show that cortical neurons that are selective for particular objects can “recognize” multiple objects in parallel at levels as high as the inferior temporal cortex. When the scene is cluttered, this initial parallel process is followed within 150 ms by competitive inhibition of all but the one relevant object in a given receptive field (e.g., Chelazzi, Duncan, Miller, & Desimone, 1998; see Rousselet, Thorpe, & Fabre-Thorpe, 2004a, for a review). The large and overlapping receptive fields found in the inferior temporal cortex would allow for detection of a target among several nontargets in parallel, followed by competitive suppression of nontargets.

Correspondence concerning this article should be addressed to Mary C. Potter, 46-4125, Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology, 77 Massachusetts Avenue, Cambridge, MA 02139. E-mail: E-mail: mpotter@mit.edu.
Laura F. Fox is now at Yale Law School

If a similar processing sequence occurs in human vision, that could account for our capacity to detect a target among multiple pictures rapidly with little interference from nontarget pictures. The subsequent zeroing in on a single item for continued processing is consistent with evidence for serial processing of individual items when the task requires it. As Rousselet et al. (2004a) said, "Constraints considerably limit the amount of information that can be processed and explicitly accessed at once, so that serial selection of objects is often necessary" (p. 369). Memory consolidation of a picture has been shown to require a much longer exposure duration than detection (e.g., Potter, 1976), suggesting that serial processing may be required for later memory for simultaneously presented pictures.

Rapid Comprehension of the Gist of a Picture

Categorical targets such as animals or vehicles can be detected accurately when a picture that the viewer has not seen before is presented as briefly as 20 ms (with no mask), and measures of event-related potential show that targets begin to be discriminated from nontargets as early as 150 ms after presentation (e.g., Thorpe, Fize, & Marlot, 1996). The gist of a scene—its main topic—can usually be reported correctly with an exposure duration of about 100 ms, followed by a mask (Davenport & Potter, 2004; Fei-Fei, Iyer, Koch, & Perona, 2007).

Although a mask interrupts perceptual processing, it does not necessarily terminate conceptual processing (Intraub, 1984; Loftus & Ginn, 1984; Loschky et al., 2007; Potter, 1976). A more effective way to control the amount of time available for processing is rapid serial visual presentation (RSVP), in which each frame masks the preceding one and presents new material to be processed. Under those conditions, processing is limited to the frame duration.[1] Studies of RSVP search using a sequence of single pictures have shown that detection of a target picture designated by a verbal title such as *picnic* or *woman on phone* is well above chance at a per-picture duration of about 113 ms (Intraub, 1981; Potter, 1975, 1976; see also Evans & Treisman, 2005).

Does Detection of a Target Picture Require Attention?

Evans and Treisman (2005) presented a sequence of pictures in RSVP that included two targets, pictures of animals or vehicles. They found that the second target was subject to an attentional blink when presented within about 500 ms of the first, indicating that detection does require attention. A different method for removing attention gave a different result, however. Li, VanRullen, Koch, and Perona (2002) used a demanding foveal task to show that a peripheral target picture (an animal or vehicle) could be detected as accurately when participants were giving primary attention to the foveal task as when they were merely fixating at that location. A second study with the same dual task (Fei-Fei, VanRullen, Koch, & Perona, 2005)[2] found that detection was as easy with two as with one peripheral picture, regardless of the distance between the two pictures. A subsequent study, however, found that when participants had to detect the contents of both pictures (rather that simply saying whether there was an animal picture or not), performance was worse when the pictures were relatively close (a center-to-center separation of 3 degrees of visual angle) than when they were separated by 8 degrees (VanRullen, Reddy, & Fei-Fei, 2005). The authors suggest that preattentive identification, although it occurs without attention, is interfered with when there is more than one stimulus in the relevant receptive field.

[1]Under some circumstances the viewer may give unequal attention to successive frames (Intraub, 1984), but when no special instructions are given, attention switches to each successive frame as it appears (Potter & Levy, 1969).
[2]F. F. Li is now L. Fei-Fei.

Target Detection Among Simultaneous Pictures

In the studies just reviewed, evidence for parallel scene or face detection was obtained in some conditions (particularly when the distractor and target were far apart), even though attention was focused on a difficult foveal task. What is the evidence for parallel processing in detection tasks with full attention? Rousselet, Fabre-Thorpe, and Thorpe (2002) presented either one or two photographs of natural scenes for 20 ms (followed by a blank screen) and participants made a go/no-go decision as to whether there was an animal pictured in the display. Go responses were equally fast whether one picture or two pictures were presented, and event-related potentials to go trials were differentiated from those to no-go trials beginning 150 ms after presentation, whether one or two pictures were presented. In a further study measuring eye saccades to the animal picture when two pictures were presented simultaneously, correct responses began to exceed errors at latencies as short as 120 ms (Kirchner & Thorpe, 2006). Thus, detection of a target seems to occur as quickly with two potential targets as with one, consistent with the results for discrimination of a target among two pictures without attention (Fei-Fei et al., 2005).

In another animal-detection study (Rousselet et al., 2004b), up to four pictures were presented simultaneously for 26 ms, and accuracy dropped somewhat as the number increased. The authors showed that most of the drop in accuracy could be accounted for if parallel and independent processing of the pictures was assumed, converging on a single output system. In their model, the probability that the target picture is missed and the increasing possibility of a false alarm as the total number of distractors increases together account for most of the drop in accuracy. A detection study by VanRullen, Reddy, and Koch (2004), in which up to 16 pictures were presented simultaneously, led to a different conclusion. As in the Rousselet et al. (2002, 2004b) experiments, the task was to detect the presence of a picture containing an animal. In a go/no-go condition the picture array was presented for 200 ms, followed by a noise mask. Accuracy dropped markedly as the number of pictures increased, falling to near chance with more than 8 pictures. For a second group of participants the array remained in view until the subject responded. Response times increased as the number of pictures in the array increased, with a slope of 40 ms per picture. These results suggest that the pictures were searched serially rather than in parallel.

Collectively, these studies reach somewhat different conclusions about whether multiple scenes can be processed in parallel. Several differences in method among these experiments could be responsible for this difference. In Rousselet et al. (2004b) the pictures were adjacent to the fixation point and were presented for 20 ms with no mask. The pictures in VanRullen et al. (2004) were much smaller and were presented for longer: either for 200 ms plus a mask or until a response was made. Under these conditions participants may have chosen to scan the pictures serially rather than responding on the basis of information in the initial short glimpse.

A different explanation for the divergence between the two studies was suggested by VanRullen et al. (2005; see also Reddy & VanRullen, 2007), based on their finding that picture detection is impaired when the distance in visual angle between pictures is small (e.g., 3 degrees or less) and (by hypothesis) pictures are likely to fall within the same receptive field in the inferior temporal cortex. In VanRullen et al. (2004) the average distance between pictures decreased as the number of pictures increased in their 4 × 4 array (15 × 10 degrees overall), consistent with this hypothesis. In Rousselet et al. (2004b) pictures were separated by only 2 degrees, and yet performance declined little between one and four pictures. Because the pictures were larger in this study, however, the distance from center to center of adjacent pictures was about 10 degrees horizontally and 8 degrees vertically, sufficient to keep a large part of each picture in separate receptive fields.

One limitation of the experiments on multiple pictures just reviewed is that they presented pictures either briefly with no mask or with a meaningless mask. As noted earlier, processing may continue after the stimulus array has terminated, even when it is followed by a meaningless mask; a meaningful, recognizable mask more effectively interrupts processing (e.g., Loschky et al., 2007). In the present study RSVP was used to ensure that pictures were followed by new, meaningful arrays.

Time to Consolidate Memory for a Scene

Subsequent recognition memory for a briefly presented picture is very good if it is presented for about 100 ms and followed by a mask or a blank screen. However, if the picture is presented in an RSVP sequence of other to-be-remembered pictures, an uninterrupted period of consolidation of up to 1,000 ms may be required to reach that same high level of recognition memory (e.g., Intraub, 1979, 1980; Potter, 1976; Potter & Levy, 1969; Potter, Staub, & O'Connor, 2004; Potter, Staub, Rado, & O'Connor, 2002). Such studies have shown that although observers can understand the gist of a novel pictured scene in a glimpse as short as 100 ms, the picture is likely to be quickly forgotten if another to-be-attended picture follows shortly. Can multiple pictures in one frame be consolidated as rapidly as a single picture?

The Present Study

Here we used a combination of RSVP and multiple simultaneous pictures to address the question of whether more than one picture can be processed simultaneously. A diagram of a trial in Experiment 1 is shown in Figure 1. Each frame in the RSVP sequence contains from none to four pictures; quadrants without pictures contain noise masks consisting of cut-up pictures. We used two different tasks, one a detection task (Experiments 1 and 3) and the other a memory task (Experiments 2 and 4). Although target detection has been used as a benchmark of successful processing in many recent studies, picture memory is an equally important measure of the success of picture processing. Memory is generally excellent when individual pictures are viewed for 2 s or more (Nickerson, 1965;Shepard, 1967;Standing, 1973). However, recognition memory drops when pictures are presented in RSVP for a duration of 500 ms or less, declining to near chance recognition at a duration of about 113 ms (Potter & Levy, 1969), even though target detection remains well above chance (Potter, 1976).

The disparity in time course between detection and memory consolidation suggests that pictures are understood rapidly but may then be quickly forgotten. Can two or more simultaneously presented pictures in an RSVP stream be consolidated at the same time as readily as a single picture? That question is addressed in Experiments 2 and 4.

Experiment 1: Detection

A single target category (animal or vehicle) was used in the work of Rousselet et al. (2002, 2004b) and VanRullen et al. (2004). Participants had extensive practice with that category, although no pictures were repeated in the experiment. In the present experiments, a new search category was specified on each RSVP trial, and the categories were often complex, such as "hands holding decorated eggs" or "cut-up fruit." This method had previously been used with RSVP sequences of single pictures (Intraub, 1981;Potter, 1975,1976). In Experiment 1 we extended this method by presenting RSVP sequences of multiple-picture frames (see Figure 1).

Method

Participants—Twenty-four volunteers (9 men, 15 women) from the Massachusetts Institute of Technology (MIT) community gave written informed consent and were paid for their

participation. Four additional participants were replaced: 3 because their false *yes* rates were over 25%, and 1 because the correct *yes* rate was below 25%. In this and later experiments, participants reported normal or corrected-to-normal vision.

Procedure—Figure 1 illustrates a trial in Experiment 1. At the start of the trial, participants saw a short descriptive title of a picture (e.g., *fox* or *people at computer*) for 2 s. Then they viewed an eight-frame RSVP sequence in which each frame consisted of four picture locations (quadrants) and contained 0, 1, 2, 3, or 4 pictures. Each sequence included a total of 8 pictures. The quadrants in a given frame that were not occupied by pictures were filled with copies of a visual mask. The first and last frames contained only masks, as did some of the other frames. A white fixation cross appeared in the center of the screen for 500 ms between the presentation of the title and the beginning of the picture sequence and remained on the screen throughout the presentation. The presentation duration on a given trial was 240, 400, or 720 ms per frame, counterbalanced across trials. Participants were instructed to maintain fixation on the cross. Their task was to press a key marked *yes* if they saw a picture somewhere in the sequence that fit the title presented at the start of the trial and press a key marked *no* if they did not.

Stimuli—All four experiments used pictures from a set of 1,152 color photographs with widely varied content, chosen from commercially available compact discs and other sources. They included pictures of animals, people engaged in various activities, landscapes, interiors, food, and city scenes; the intent was to sample as wide a range of pictures as possible. Pictures were assigned randomly to each sequence, except that pictures similar in subject matter were not included in the same sequence. The masks used were 16 different texture images made by fragmenting 300 × 200 pixel photographs into 10 × 10 pixel squares and randomly reassembling the pieces; the photographs were from the same set (but not the same pictures) as the experimental pictures. The masks on a given frame were identical, but the masks were different for each frame in a sequence. Both pictures and masks were stored as 300 × 200 pixel JPEG files. Images were displayed 15 pixels above or 15 pixels below the vertical center of the screen and 15 pixels to the right or 15 pixels to the left of the horizontal center of the screen. The images on a frame together subtended about 14.5 degrees of visual angle vertically and about 20 degrees horizontally. No pictures were repeated.

Brief verbal descriptions were written that captured the meaning or gist of each picture. The descriptive titles did not include specific color or shape information. They ranged in length and complexity from simple one-word names (e.g., *moose* or *bicycle*) to longer, more specific phrases (e.g., *people washing hands in stream* or *ornate old building with fountain*). A longer description was used only when it seemed necessary to convey the gist of a picture.

Apparatus—All experiments were run using Matlab 5.2.1 on a PowerMac G3 with an Apple 17-in. (43.18-cm) studio display. The screen was set to 1,024 × 768 resolution with a 75 Hz vertical refresh rate and 32-bit colors. The testing room was normally illuminated.

Design—Seventy-two RSVP sequences included the named target (target-present trials), and 36 sequences did not include the target (target-absent trials). Titles for pictures that did not appear in the experiment were assigned randomly to the target-absent trials. Across trials, an equal number of pictures were presented in frames with 1, 2, 3, or 4 pictures (e.g., there were four times as many frames with 1 picture as with 4 pictures). The target (on target-present trials) was equally likely to appear alone or with 1, 2, or 3 distractor pictures. The target was equally likely to be in serial positions 2-7 (serial positions 1 and 8 consisted only of masks). The number of pictures with the target was counterbalanced with frame duration, within and between subjects. Within subjects, the presentation quadrant of a target was also counter-balanced. The order of trials was randomized and the resulting order was constant for all participants.

To sample from a wide range of the pictures, we randomly chose two possible target pictures for each trial, constrained by the necessity to counterbalance presentation duration, number of pictures on the target frame, and quadrant in which the target appeared. Half the participants searched for one of these pictures and half for the other.

Results and Discussion

Figure 2 shows the main results. Target detection was very good; overall, 76% of the target pictures were detected on target-present trials, with 8% false alarms on target-absent trials. Even in the most difficult condition, with four pictures on the target frame and presentation for 240 ms per frame, the target was detected on 59% of the trials, with a false alarm rate of 9%. Overall, the more pictures presented simultaneously with a given picture, the lower the detection rate. As expected, detection was better at longer presentation durations.

To look at the effects of the main variables on correct detection, an analysis of variance (ANOVA) was performed on *yes* responses in target-present trials only, with number of pictures on the target frame and presentation duration as within-subject variables. The analysis revealed an effect of number of pictures, $F(3, 69) = 8.86$, $p < .001$, with higher detection rates for targets presented among fewer distractor pictures. A Newman-Keuls test of the differences between the number-of-pictures means showed that the difference between one picture with no distractors ($M = 84\%$) and each of the other numbers of pictures (76%, 71%, and 71% for two, three, and four pictures, respectively) was significant, $q(2, 69) = 3.81$, $q(3, 69) = 6.24$, and $q(4, 69) = 6.36$, $p < .01$ in each case. There were no significant differences between two, three, and four pictures. There was also a main effect of presentation duration, $F(2, 46) = 56.20$, $p < .001$, with means of 64%, 77%, and 86% for durations of 240, 400, and 720 ms, respectively. The interaction between number and duration was not significant, $F(6, 138) = 1.70$, $p = .13$.

To sum up the results of Experiment 1, detection of a target picture was surprisingly good. When the target was presented alone, detection accuracy overall was higher (84%) than when there were also one (76%) or more (71%) distractors, although the number of distractors (one, two, or three) did not make a significant difference. Thus, although detection was good even with three distractors on the frame, the result does not support an unlimited-capacity, parallel model of detection. We return to this point in the General Discussion.

Experiment 2: Recognition Memory

Target detection is only one measure of successful picture processing. The consolidation of a picture into short-term memory is an equally important part of processing. Studies of visual short-term memory for objects such as colored geometric figures have shown a capacity for retention of about four objects (e.g., Luck & Vogel, 1997). In recent studies, Potter et al. (2002, 2004) have shown that a pictured scene presented for 173 ms in RSVP can be remembered fairly accurately when tested immediately. Memory for multiple simultaneous pictures has not been tested previously, however.

Method

Participants—Eighteen volunteers (6 men, 12 women) from the MIT community gave written informed consent and were paid. None had participated in Experiment 1. Two additional participants were replaced because they had unacceptably high false alarm rates (greater than 25%).

Procedure—We designed Experiment 2 to be similar to Experiment 1, despite the different task. As in Experiment 1, on each trial participants viewed an eight-frame RSVP sequence

including a total of eight different pictures, with zero to four pictures on a given frame. Again, presentation duration per frame was 240, 400, or 720 ms.

We asked participants to view and remember the pictures. Unlike Experiment 1, participants did not see a title and did not have a detection task. A *yes-no* recognition test began 200 ms after the RSVP sequence, consisting of four pictures from the presentation sequence interspersed with four new distractor pictures. Test pictures were presented one at a time in the center of the screen for 400 ms followed by a blank screen until the subject responded by pressing a labeled key on the keyboard. Participants were instructed to make a *yes* response if they recognized a tested picture as having been in the presentation sequence and a *no* response otherwise. No feedback was given.

Stimuli—Experiment 2 employed the same pictures and masks as Experiment 1.

Design—There were 72 trials plus 3 practice trials. As in Experiment 1, the number of pictures on a given frame varied from one to four (plus some frames with no pictures, only masks), and an equal number of pictures appeared in each of these conditions, across trials. Pictures were equally likely to appear in each of the four picture quadrants and in each of the six serial presentation positions (excluding the first and last frames, which contained only masks). Pictures from a given number-of-pictures condition were equally likely to be tested in each of the eight test positions. The order in which pictures were tested was random relative to their order of presentation. As in Experiment 1, three different presentation durations were used (240, 400, and 720 ms), counterbalanced within and between subjects. On a given trial, four pictures were presented and tested, four pictures were presented but not tested, and four additional pictures served as distractors in the test. Which particular set of four pictures fulfilled each of these three roles was counterbalanced between subjects, so that a given picture served equally often as an “old” picture and as a distractor.

Results and Discussion

Figure 3 shows the main results. Experiment 2’s results contrast with those of Experiment 1 in that recognition performance in Experiment 2 was much lower than detection in Experiment 1 (overall, 38% correct recognition of old pictures and a false alarm rate of 14% for new, distractor pictures, compared with 76% and 8%, respectively, in Experiment 1) and the negative effect of increasing the number of simultaneous pictures was somewhat greater (a *d*’ comparison is given in the Appendix).

An ANOVA of *yes* responses to old pictures, with number of pictures, presentation duration, and test position as within-subjects variables, found significant main effects of each of the variables. For 1, 2, 3, or 4 pictures per frame the mean percentages of correct responses were 54%, 38%, 31%, and 31%, respectively, $F(3, 51) = 37.08$, $p < .001$. A Newman-Keuls test showed that all the means differed from each other at the .05 level or better, except that the three- and four-picture conditions did not differ. For presentation durations of 240, 400, and 720 ms per frame, the mean percentages of correct responses were 32%, 38%, and 45%, respectively, $F(2, 34) = 46.99$, $p < .001$. Presentation duration interacted with number of pictures, $F(6, 102) = 3.27$, $p < .01$; as Figure 3 shows, the benefit of a longer viewing time was greater when the tested picture was presented alone or with only one other picture.

For test position (which was randomized with respect to the order of presentation of the old pictures), the percentage of *yes* responses to old pictures declined in the course of the eight-item recognition test from 63% to 30% (reaching an asymptote at the fifth test item), $F(7, 119) = 28.23$, $p < .001$; this is consistent with recent research showing that memory for rapidly presented pictures is not lost instantly but over a period of several seconds after viewing (Potter et al., 2002, 2004). Number of pictures interacted with test position, $F(21, 357) = 1.99$, $p <$.

01, although there was no evident pattern to this effect. Finally, the three-way interaction among duration, number, and test position was also significant, $F(42, 714) = 1.47$, $p < .05$, with no apparent pattern.

To sum up the main results of Experiment 2, memory was best for pictures presented alone and became increasingly worse as more pictures were added on the same frame, reaching asymptote at three pictures per frame. Thus, multiple pictures could not be consolidated at the same time without a cost. Memory improved with increasing presentation time, particularly when the picture was presented alone.

A final question is whether recognition memory was underestimated because the spatial position and context of the picture were changed at test.[3] In a pilot study ($N = 6$) using a presentation duration of 400 ms/frame, picture memory was tested by presenting the test picture in its original frame or (on negative trials) replacing that picture with a new picture in that same frame. The to-be-recognized picture was indicated by a symbol next to its outer corner; only one picture (new or old) was tested per trial. In a control group ($N = 6$), the test picture was presented alone, as in Experiment 2. There was no difference between the two groups in overall accuracy (combining old and new trials, $M = 0.67$ for the context group, and $M = 0.66$ for the control group), although the group with context was more strongly biased to say yes, with a higher false alarm rate balanced by a higher hit rate. The results were also very close to those in Experiment 2 for the first-tested picture, at the same duration. Thus, there was no support for the idea that recognition performance was reduced in Experiment 2 because of the switch from a four-quadrant frame during presentation to a single, centered picture at test.

Summary

Although detection of a target picture in Experiment 1 was much more accurate than recognition memory for pictures in Experiment 2, the effects of number of pictures on a frame and the duration of presentation were similar in the two experiments. With respect to our main question about the ability to process multiple pictures presented simultaneously, there was a drop in performance in both tasks as the number of pictures on a frame increased. A two-stage model that accommodates these findings is presented in the General Discussion.

Rationale for Experiments 3 and 4

In Experiments 3 and 4 we evaluated the ability to detect or remember a picture when each frame is presented for only 160 ms, followed by a variable blank interstimulus interval (ISI). That brief presentation time is too short for a planned eye movement, whereas in the two longer durations in Experiments 1 and 2, participants could have moved their eyes and made one or two fixations on individual pictures, even though the instructions asked them to keep their eyes on the fixation cross. The blank ISI in Experiments 3 and 4 was 80, 240, or 560 ms, creating a stimulus onset asynchrony (SOA) of 240, 400, or 720 ms. Thus, participants in Experiments 3 and 4 had the same total time as participants in Experiments 1 and 2 to process the pictures before the next frame appeared, but they could not gain extra information by moving their eyes or continuing to observe the pictures.

In studies using an RSVP stream of single pictures, Intraub (1980) presented pictures for 110 ms with blank ISIs ranging from 0 to 5.9 s. She found that memory performance improved markedly as the ISI increased, from 20% correct at 0 ISI to an asymptote of 84% correct at an ISI of 1.5 s (see also Potter, 1976, Experiment 3, and Potter et al., 2004, Experiment 3). This finding indicates that most of the stimulus information required for processing a picture to the level required for recognition of the picture a few moments later is provided by the first 110

[3]This possibility was suggested by a reviewer.

ms of presentation plus visual persistence (iconic memory).[4] The Intraub study shows that the main benefit of a still longer presentation time is that it permits consolidation of this information, not that it allows the viewer to continue to pick up further information. That study, however, involved presentations of RSVP sequences of single pictures, not multiple simultaneous pictures.

If information from up to four simultaneous pictures is extracted as rapidly as information from one picture, then presentation with a blank ISI might be as useful for target detection and memory as continued viewing of the array.

Experiment 3: Detection With a Blank ISI

Experiment 3 replicated Experiment 1 precisely except that the frames were presented for 160 ms, followed by a variable ISI that resulted in SOAs of 240, 400, and 720 ms, as in Experiment 1.

Method

Participants—Twenty-four volunteers (9 men, 15 women) from the MIT community gave written informed consent and were paid for their participation. None had participated in Experiment 1 or 2.

Procedure and design—These were identical to those of Experiment 1, with the exception that the frames in the presentation sequence did not remain on the screen for the full presentation duration. Rather, each frame remained on the screen for only 160 ms, followed by a plain black screen (with fixation cross) for the rest of the original presentation duration. Thus, the "presentation duration" variable became the SOA in this experiment.

Results and Discussion

Figure 4 shows the main results. As in Experiment 1, overall detection performance in Experiment 3 was very good, above 60% accuracy even at an SOA of 240 ms. The number of pictures on the target frame and the SOA both had significant effects on detection accuracy.

An ANOVA of *yes* responses on target-present trials only, with number of pictures and SOA as variables, revealed a significant effect of number of pictures, $F(3, 69) = 7.06$, $p < .001$. Planned comparisons using the Newman-Keuls test showed that only the difference between 4 pictures ($M = 62\%$) and 1 (75%), 2 (69%), or 3 pictures (72%) was significant; the latter three did not differ. The effect of SOA was also significant, $F(2, 46) = 10.33$, $p < .001$, with better performance at longer SOAs (62%, 70%, and 75% for SOAs of 240, 400, and 720 ms, respectively). There was no interaction between SOA and number of pictures, $F < 1.0$.

An ANOVA comparing *yes* responses to target-present trials in Experiments 1 and 3, with experiment, number of pictures, and presentation duration/SOA as variables, showed that performance was significantly better in Experiment 1 (76%) than in Experiment 3 with a blank ISI (69%), $F(1, 46) = 8.36$, $p < .01$. There was a main effect of number of pictures, $F(3, 138) = 13.48$, $p < .001$, and a main effect of presentation duration/SOA, $F(2, 92) = 48.95$, $p < .001$. Duration/SOA interacted with experiment, $F(2, 92) = 3.27$, $p < .05$: The effect of SOA in Experiment 3 was smaller than the effect of presentation duration in Experiment 1. No other interactions were significant.

[4]Under photopic viewing conditions like those in the present experiments, iconic memory might add about 100 ms.

In sum, target detection was very good in all conditions, even when participants had as little as 160 ms (plus an ISI of 80 ms) to view a frame with four pictures. In this extreme condition, participants made a correct detection on 60% of the trials, compared with 8% false *yes* responses to no-target trials at the same SOA. The combined results from Experiments 1 and 3 show that increasing the number of pictures competing with the target on the same frame from none to three does decrease performance from 79% to 66% (on average), showing that detection among multiple simultaneous pictures is not cost free. What is surprising is that detection is so good even with three competitors. How participants are able to detect a target so accurately in the presence of competing pictures is considered in the General Discussion.

Experiment 4: Recognition With a Blank ISI

Experiment 4 was a replication of Experiment 2, but as in Experiment 3 each frame of the RSVP sequence was presented for only 160 ms, followed by a blank screen for an ISI equivalent to the remainder of the original presentation duration. We asked whether the information picked up in the first 160 ms of processing would be sufficient to permit memory consolidation, provided that additional blank time followed the presentation. As discussed earlier, Intraub (1980) found that providing blank time after a 110 ms presentation was almost as useful to later memory as giving full viewing time when a single picture was presented on each RSVP frame. We asked whether the same would be true when multiple pictures appear on a frame.

Method

The method was like that of Experiment 2, except as specified.

Participants—Eighteen volunteers (11 men, 7 women) from the MIT community gave written informed consent and were paid for their participation. All were 18-35 years old and had normal or corrected-to-normal vision. None had participated in Experiments 1-3.

Design and procedure—These were identical to those of Experiment 2, with the exception that the frames in the presentation sequence did not remain on the screen for the entire SOA. As in Experiment 3, each frame remained on the screen for only 160 ms, followed by a plain black screen (with fixation cross) for a variable ISI. The SOAs between frames were the same as the presentation durations in Experiment 2.

Results and Discussion

Figure 5 shows the main results of Experiment 4. The pattern of results was similar to that of Experiment 2, although the benefits of fewer pictures per frame and of longer SOAs were less marked.

An ANOVA of *yes* responses to old pictures only, with number of pictures, SOA, and test position as variables, showed a main effect of number of pictures, $F(3, 51) = 16.35$, $p < .001$. A Newman-Keuls test showed a significant difference between having just 1 picture (43%) and having 2 (34%), 3 (33%), or 4 pictures (31%), $p < .01$, although the latter three did not differ. There was also a main effect of SOA, $F(2, 34) = 6.80$, $p < .01$; 31%, 36%, and 38% of pictures were recognized at SOAs of 240, 400, and 720 ms, respectively. Number and SOA interacted, $F(6, 102) = 2.74$, $p < .05$, with a larger effect of increasing the SOA when there were fewer pictures; indeed, with three or four pictures there was no overall benefit of a longer SOA. Test position was also significant, $F(7, 119) = 24.43$, $p < .001$; performance dropped from 58% to 29% across test positions 1-8. SOA and test position interacted, $F(14, 238) = 1.97$, $p < .05$, with no clear pattern. There were no other interactions.

An analysis comparing the results of Experiments 2 and 4 was performed, to examine the effects of the ISI manipulation. In an ANOVA of correct *yes* responses with experiment, duration (called SOA in Experiment 4), test position, and number of pictures as the variables, there was no significant difference between the experiments, $F < 1.0$. There was a main effect of number of pictures, $F(3, 102) = 52.92$, $p < .001$; a main effect of duration/SOA, $F(2, 68) = 37.63$, $p < .001$; and a main effect of test position, $F(7, 238) = 54.07$, $p < .001$. There was an interaction between number of pictures and duration/SOA, $F(6, 204) = 5.63$, $p < .001$, with duration/SOA having a larger effect when there were fewer pictures. There were two significant interactions with experiment: Number of pictures had a larger effect in Experiment 2, $F(3, 102) = 52.92$, $p < .001$; and so did SOA/duration, $F(2, 68) = 4.45$, $p < .05$.

In summary, replacing some of the picture presentation time with an unfilled ISI reduced not only the differential effect of duration/SOA on picture memory, but also the effect of the number of pictures presented simultaneously. The general pattern of results, however, was the same as that observed in Experiment 2.

Just as recognition memory in Experiment 2 was worse than detection in Experiment 1, recognition in Experiment 4 was worse than detection in Experiment 3. A *d*' analysis comparing Experiments 3 and 4 is given in the Appendix.

Summary

The results of Experiments 2 and 4, taken together, indicate that viewers consolidate single pictures into memory much more easily than they consolidate multiple simultaneous pictures. As a result, recognition performance decreases as more pictures are presented. Added time per frame (especially when the pictures remain in view) improves picture memory, as shown previously (Intraub, 1980; Potter, 1976; Potter & Levy, 1969), but the improvement is greater the fewer the pictures presented on a frame. There was, in short, no support for the hypothesis that simultaneously presented pictures are consolidated in memory in parallel without mutual interference.

General Discussion

The main results of these experiments may be summarized as follows. Both detection and recognition memory were more accurate when fewer pictures were presented simultaneously. Thus, to answer to our original question, in neither task could viewers process multiple pictures cost free, contrary to some previous claims (e.g., Rousselet et al., 2004b). The absolute effect of the number of pictures was similar in magnitude in the two tasks, but because recognition memory was much less accurate than target detection, the proportional impact was much greater for recognition than for detection. The beneficial effect of increasing the presentation duration (or SOA) was also found in both tasks, with a somewhat different pattern: For the recognition task, increasing the presentation duration had a larger effect when there was only one picture on the frame (and all the extra time could be directed to that one picture), whereas with detection the benefit of a longer duration appeared to have the least effect when there was just one picture (because in most cases that picture had already been detected, even at the shortest duration).

Gist Detection in a First Pass?

Why was detection performance so good in Experiments 1 and 3, even though pictures were never repeated, there was no training in detecting particular categories, and category information was provided only seconds before the very rapid presentation? One hypothesis is that the target titles were usually a good fit to the gist of the target picture, and gist is what viewers normally extract first. Whether the gist of as many as four pictures can be extracted

in parallel is unclear, however. Did participants instead do a rapid feature search based on the title, as suggested by Evans and Treisman (2005)? The search titles gave specific conceptual information (e.g., *hands holding decorated eggs*) but did not specify low-level features except by implication: The eggs would be egg-shaped and would not be plain white, there would be two (perhaps more) hand-shaped objects grasping or exhibiting the eggs, and the scale would be appropriately close-up. However, even these features do not seem specific enough for a strictly bottom-up, feature-based search process, although they might be sufficient to pick out the most likely candidate picture in a multipicture frame that would then be checked by focal attention. The low false *yes* rate in all conditions rules out the sole use of weak probabilistic features in detection.

A Two-Stage Model of Processing in Detection and Memory Tasks

In this section we suggest a simple two-stage model to account for the present results. A brief, parallel stage of processing (usually sufficient for detection but rarely for later memory) is followed by a serial second stage in which attention is directed to one picture at a time (allowing for confirmation of detection and for memory consolidation). Two-stage models of visual processing have been proposed by many theorists, particularly for search tasks (e.g., Bundesen, 1990). One pertinent example is Wolfe's (1994) guided search model of search for a target in a single array. We propose a model in the spirit of the Wolfe model, which applies sequentially to each frame, beginning anew with the next frame. Stage 1 consists of a rapid global process that takes information from the whole frame and selects one picture for focal attention in Stage 2. This first pass may be serial but very rapid; functionally, however, it has a parallel component in that all the pictures compete for selection. In the detection task (Experiments 1 and 3) this initial stage of processing allows the viewer to select a likely target candidate; in the memory task (Experiments 2 and 4), the initial pass results in the selection of one picture randomly or on the basis of relative salience.

In both tasks, the selected picture is processed further in the second, serial stage, as in Wolfe's (1994) guided-search model (see also Chun & Potter, 1995, and Treisman & Gelade, 1980). In the detection task, the selected picture is confirmed if it fits the target specification; if it does not fit the specification, attention shifts to another picture as long as the array remains in view. In the memory task, the selected picture continues to be processed until it is identified and consolidated; attention then moves to another picture if the array remains in view. Because consolidation may take longer than the frame duration, frequently only one picture per frame receives Stage 2 processing.

According to this simple model, the results of Experiments 1 and 3 indicate that Stage 1 processing is sufficient for correct detection of a specified target on more than half the trials even with four pictures and a duration of 240 ms (or 160 ms plus a brief ISI). With just one picture on the frame, it is the one selected, and performance is near ceiling; extra time is helpful only if confirming the picture's match to the target description happens to be difficult. With more than one picture, adding viewing time gives a chance for attention to switch to a second potential target if the first choice was mistaken and if there is another picture on the frame.

In the recognition memory task in Experiments 2 and 4, Stage 1 processing again provides a quick overview of all pictures on the frame. One picture is again selected for Stage 2 processing, perhaps on the basis of bottom-up salience. Because consolidation often takes longer than the longest duration in the present studies (720 ms), the benefit of extra viewing time is most concentrated when just one picture is presented on a frame and is diluted when more than one picture is presented. On the other hand, recognition memory when there are as many as three or four pictures remains significantly above the false alarm rate, even with a presentation duration of 160 or 240 ms, suggesting that the first stage of processing increases the probability of remembering all pictures to some degree.[5]

The model gives a good account of the effects of the number of pictures on a frame. In the recognition task, the probability that a given picture has been selected for focal attention decreases as expected with the number of pictures, and that is the major determinant of whether the picture is recognized later; the hypothesized Stage 1 process accounts for only a small proportion of correct recognitions. In the detection task, although the number of pictures on a frame also affects the likelihood of correct detection, detection is quite high in all conditions. If performance at an SOA of 240 ms with four pictures on the frame is taken as an approximate measure of successful Stage 1 processing, then the target is detected correctly in Stage 1 on about 50% of the trials (after subtracting false alarms).

The model assumes that longer exposure durations affect only the second stage of focal attention. In the case of recognition memory, the extra time is most effective when there is only one picture on the frame because even a single picture has often not reached asymptote by 720 ms, the longest duration in the present experiments. In the detection task, in contrast, extended time in Stage 2 is least important when the target picture is alone, because it has been selected and usually identified in Stage 1. When there is more than one picture, Stage 2 permits shifts in attention that increase correct detection. Increasing the blank ISI after a presentation of 160 ms in Experiments 3 and 4 also improved performance a little in both tasks, but not nearly as much as did increasing the presentation duration in Experiments 1 and 2.

Can More Than One Picture Be Processed at the Same Time?

As already noted, the results clearly indicate that processing more than one picture at the same time is not cost-free as some have suggested. When the task was to detect a picture consistent with a verbal title (Experiments 1 and 3), however, much of the relevant processing appeared to occur in a first, possibly parallel stage, encompassing all four pictures and taking less than 240 ms. When the task was to remember pictures (Experiments 2 and 4), processing of the same picture arrays appeared to be more nearly serial, fit well by a model that assumes that one picture was chosen in the first stage for focal attention in the second, serial stage, giving a singleton picture a marked advantage over multiple pictures on the same frame.

Other Evidence for a Two-Stage Model

Recent work has shown that an eye movement to the target picture of an animal can be initiated as early as 120 ms after two simultaneous pictures have been presented (Kirchner & Thorpe, 2006; Rousselet et al., 2002). Because planning and executing a command to move the eyes takes a minimum of about 70 ms, this finding suggests that attention is attracted to a target as little as 50 ms after the onset of two pictures. This result is consistent with the present claim that targets can be provisionally detected very rapidly.

Our results show that there is competition among the pictures on a frame before attention becomes focused on the target, because target detection in Experiment 1 dropped significantly as the number of pictures on the frame with the target increased, particularly at the shortest

[5]Recognition memory would be expected to be above chance because, even with four pictures on the frame, there is a .25 probability that the one selected is the one tested for recognition. If we take the proportion of single pictures that are remembered correctly (minus the false alarm rate) as an estimate of the likelihood that the selected picture will be remembered, then that proportion, divided by *N* (the number of pictures on the frame), gives the estimated above-chance proportion correct when there are *N* pictures. For example, .41 single pictures in the 240 ms condition in Experiment 2 were correctly recognized; subtracting the false alarm rate of .15, the above-chance recognition probability was .26: That is our estimate of the likelihood that if the picture was the one selected, it would be correctly recognized later. If there were four pictures on the frame and only one was selected, the average recognition probability should be .065 (.26 divided by 4). The observed above chance probability was .12, .055 above the predicted performance. In each condition the observed performance was somewhat higher (by about .06 on average) than the predicted performance, suggesting that there was some processing of pictures other than the one selected when there was more than one picture on the frame. We hypothesize that this processing occurred in the first, global stage, consistent with the observation that the above-chance effect was similar in magnitude in Experiment 4, with a presentation duration of 160 ms plus a variable ISI, to that in Experiment 2. Had the extra processing occurred in Stage 2, the effect should have been reduced in Experiment 4.

duration. Moreover, not all the targets attracted attention immediately: The percentage of correct detections increased with an increase in presentation duration.

Recent data on the monkey brain support a two-stage model of the kind we propose. Neurons in the monkey brain that are selective for an object can each respond to their individual preferred objects in parallel at levels as high as the inferior temporal cortex (for reviews see Hung, Kreiman, Poggio, & DiCarlo, 2005; Rousselet et al., 2004a). In a cluttered scene an initial parallel process is quickly followed by competitive inhibition of responses to all but one object in a given receptive field (e.g., Chelazzi et al., 1998). Such a processing sequence is just what we propose to account for the differences between detection and memory: Initial detection occurs in parallel, followed by selective attention to the most likely target. Memory consolidation, we hypothesize, begins when attention has been directed to one picture, and other pictures are inhibited until a change in focal attention occurs.

We conclude that multiple pictures will initially be processed simultaneously (or serially, but very rapidly) and subsequently one at a time, more slowly. When the task is to search for a target defined conceptually, the initial scan of up to at least four simultaneous pictures is often sufficient for detection. In contrast, when the task is to remember the pictures, slower serial memory consolidation is almost always required.

Acknowledgments

This work was supported by Grant MH47432 from the National Institute of Mental Health. We thank Virginia Valian and Jodi Davenport for their comments on an earlier version of the manuscript, and Nina Strohminger and Jennifer Olejarczyk for research assistance.

Appendix

Comparisons Between Detection and Recognition Tasks

Analyses Comparing Experiments 1 and 2

A comparison between the correct *yes* results of Experiments 1 and 2 in Figures 2 and 3 suggests that target search is much more accurate than recognition memory. However, because false *yes* rates differed in the two experiments, we carried out a *d*' analysis on individual performance (which took into account both correct and false *yes* responses). Because accuracy declined significantly over the recognition test of eight pictures (as in Potter et al., 2002,2004), whereas in the detection task there was only a single response on each trial, we based this analysis on only the first of the eight recognition tests on each trial of Experiment 2; overall, half of the first test pictures were distractors and half were old pictures, balanced over condition.

In each experiment, responses at each Presentation Duration × Number of Pictures condition were analyzed separately for each subject; there were six positive detection tests and three recognition tests per condition per subject that were used in calculating *P*(hits). The false alarm rate was separated by presentation duration but not by number of pictures, because number of pictures was meaningless in recognition tests of distractors and in target-absent detection trials (in target-absent trials, there were always eight pictures, any of which could have generated a false alarm). Thus, the same false alarm rate was used across the number-of-pictures conditions. *P*(false alarm) was based on 12 recognition tests and 12 detection trials per duration, per subject. Scores of 1.0 or 0.0 were adjusted by subtracting or adding (respectively) half the average step size.[A1]

[A1]For hits, the step size was .167 for detection and .33 for recognition; for false alarms, the step size was .083 for both detection and recognition.

An analysis of variance of the *d*' scores for Experiments 1 and 2 showed a highly significant effect of experiment, $F(1, 40) = 119.10$, $p < .001$. For detection in Experiment 1, mean *d*' was 2.11; for recognition in Experiment 2, mean *d*' was 0.87. There were also main effects of number of pictures, $F(3, 120) = 13.79$, $p < .001$; and duration, $F(2, 80) = 21.19$, $p < .001$. No interactions were significant. Thus, although numerically the effect of number of simultaneous pictures was somewhat greater in the recognition task, the size of the effect was not significantly different from that for detection.

Analyses Comparing Experiments 3 and 4

A *d*' analysis was carried out comparing detection in Experiment 3 with recognition memory in Experiment 4, like that comparing Experiments 1 and 2. As before, only the first recognition test in Experiment 4 was included, to make the recognition task more comparable to the immediacy of detection. Detection in Experiment 3, (*d*') = 1.88 was much better than recognition in Experiment 4, (*d*') = 0.76, $F(1, 40) = 110.25$, $p < .001$. The main effect of number of pictures was significant, $F(3, 120) = 8.40$, $p < .001$, but not the main effect of SOA/duration, $F(2, 80) = 2.47$, $p = .091$. The comparison between Experiments 3 and 4 was thus similar to that between Experiments 1 and 2: a significant main effect of detection versus recognition, but no interaction with other variables.

References

Bundesen C. A theory of visual attention. Psychological Review 1990;97:523–547. [PubMed: 2247540]

Chelazzi L, Duncan J, Miller EK, Desimone R. Responses of neurons in inferior temporal cortex during memory-guided visual search. Journal of Neurophysiology 1998;80:2918–2940. [PubMed: 9862896]

Chun MM, Potter MC. A two-stage model for multiple target detection in rapid serial visual presentation. Journal of Experimental Psychology: Human Perception and Performance 1995;21:109–127. [PubMed: 7707027]

Davenport JL, Potter MC. Scene consistency in object and background perception. Psychological Science 2004;15(8):559–564. [PubMed: 15271002]

Evans KK, Treisman A. Perception of objects in natural scenes: Is it really attention-free? Journal of Experimental Psychology: Human Perception and Performance 2005;31:1476–1492. [PubMed: 16366803]

Fei-Fei L, Iyer A, Koch C, Perona P. What do we perceive in a glance of a real-world scene? Journal of Vision 2007;7(1):1–29. [PubMed: 17997664]

Fei-Fei L, VanRullen R, Koch C, Perona P. Why does natural scene categorization require little attention? Exploring attentional requirements for natural and synthetic stimuli. Visual Cognition 2005;12:893–924.

Hung CP, Kreiman G, Poggio T, DiCarlo JJ. Fast readout of object identity from macaque inferior temporal cortex. Science November 4;2005 310:863–866. [PubMed: 16272124]

Intraub H. The role of implicit naming in pictorial encoding. Journal of Experimental Psychology: Human Learning and Memory 1979;5:1–12.

Intraub H. Presentation rate and the representation of briefly glimpsed pictures in memory. Journal of Experimental Psychology: Human Learning and Memory 1980;6:1–12. [PubMed: 7373241]

Intraub H. Rapid conceptual identification of sequentially presented pictures. Journal of Experimental Psychology: Human Perception and Performance 1981;7:604–610.

Intraub H. Conceptual masking: The effects of subsequent visual events on memory for pictures. Journal of Experimental Psychology: Learning, Memory, and Cognition 1984;10:115–125.

Kirchner H, Thorpe SJ. Ultra-rapid object detection with saccadic eye movements: Visual processing speed revisited. Vision Research 2006;46:1762–1776. [PubMed: 16289663]

Li FF, VanRullen R, Koch C, Perona P. Rapid natural scene categorization in the near absence of attention. Proceedings of the National Academy of Sciences 2002;99:9596–9601.

Loftus GR, Ginn M. Perceptual and conceptual masking of pictures. Journal of Experimental Psychology: Learning, Memory, and Cognition 1984;10:435–441.

Loschky LC, Sethi A, Simons DJ, Pydimarri TN, Ochs D, Corbeille JL. The importance of information localization in scene gist recognition. Journal of Experimental Psychology: Human Perception and Performance 2007;33:1431–1450. [PubMed: 18085955]

Luck SJ, Vogel EK. The capacity of visual working memory for features and conjunctions. Nature November 20;1997 390:279–281. [PubMed: 9384378]

Nickerson RS. Short-term memory for complex meaningful visual configurations: A demonstration of capacity. Canadian Journal of Psychology 1965;19:155–160. [PubMed: 14296000]

Potter MC. Meaning in visual search. Science March 14;1975 187:965–966. [PubMed: 1145183]

Potter MC. Short-term conceptual memory for pictures. Journal of Experimental Psychology: Human Learning and Memory 1976;2:509–522. [PubMed: 1003124]

Potter MC, Levy EI. Recognition memory for a rapid sequence of pictures. Journal of Experimental Psychology 1969;81:10–15. [PubMed: 5812164]

Potter MC, Staub A, O'Connor DH. Pictorial and conceptual representation of glimpsed pictures. Journal of Experimental Psychology: Human Perception and Performance 2004;30:478–489. [PubMed: 15161380]

Potter MC, Staub A, Rado J, O'Connor DH. Recognition memory for briefly presented pictures: The time course of rapid forgetting. Journal of Experimental Psychology: Human Perception and Performance 2002;28:1163–1175. [PubMed: 12421062]

Reddy L, VanRullen R. Spacing affects some but not all visual searches: Implications for theories of attention and crowding. Journal of Vision 2007;7(2):1–17. [PubMed: 18217818]

Rousselet G, Fabre-Thorpe M, Thorpe SJ. Parallel processing in high level categorization of natural images. Nature Neuroscience 2002;5:629–630.

Rousselet G, Thorpe SJ, Fabre-Thorpe M. How parallel is visual processing in the ventral pathway? Trends in Cognitive Sciences 2004a;8:363–370. [PubMed: 15335463]

Rousselet GA, Thorpe SJ, Fabre-Thorpe M. Processing of one, two or four natural scenes in humans: The limits of parallelism. Vision Research 2004b;44:877–894. [PubMed: 14992832]

Shepard RN. Recognition memory for words, sentences, and pictures. Journal of Verbal Learning and Verbal Behavior 1967;6:156–163.

Standing L. Learning 10,000 pictures. Quarterly Journal of Experimental Psychology 1973;25:207–222. [PubMed: 4515818]

Thorpe S, Fize D, Marlot C. Speed of processing in the human visual system. Nature June 6;1996 381:520–522. [PubMed: 8632824]

Treisman AM, Gelade G. A feature-integration theory of attention. Cognitive Psychology 1980;12:97–136. [PubMed: 7351125]

VanRullen R, Reddy L, Fei-Fei L. Binding is a local problem for natural objects and scenes. Vision Research 2005;45:3133–3144. [PubMed: 16023696]

VanRullen R, Reddy L, Koch C. Visual search and dual-tasks reveal two distinct attentional resources. Journal of Cognitive Neuroscience 2004;16(1):4–14. [PubMed: 15006031]

Wolfe JM. Guided Search 2.0: A revised model of visual search. Psychonomic Bulletin & Review 1994;1:202–238.

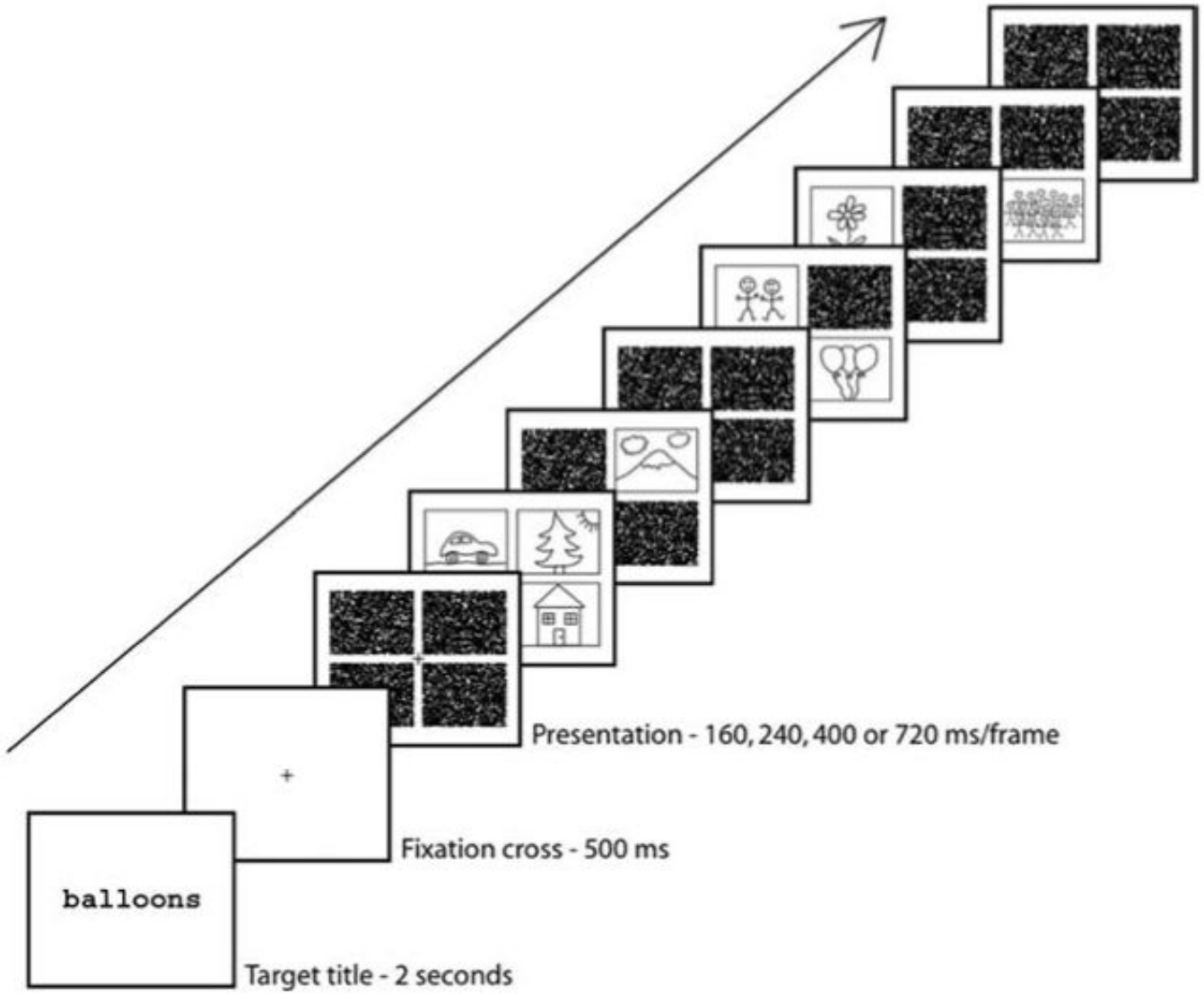

Figure 1.
Schematic representation of the rapid serial visual presentation sequence on a trial of Experiment 1. Line drawings are used in the figure for clarity; the actual experiment employed color photographs displayed on a black background.

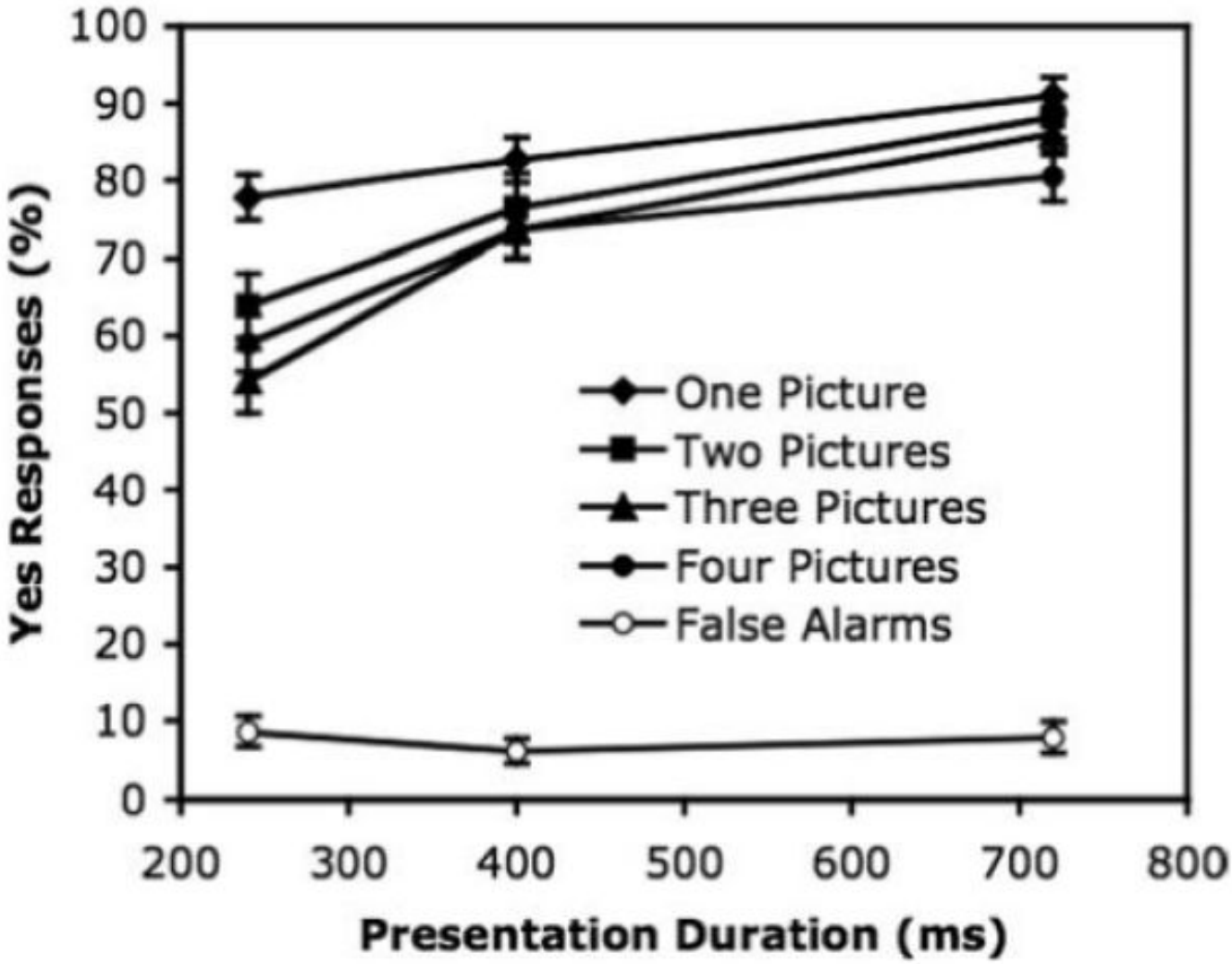

Figure 2.
Experiment 1: Percentage of correct *yes* responses to targets at each presentation duration, separately for target frames with 1, 2, 3, or 4 pictures. The percentage of false *yes* responses on target-absent sequences is also shown for each presentation duration. Error bars represent standard error.

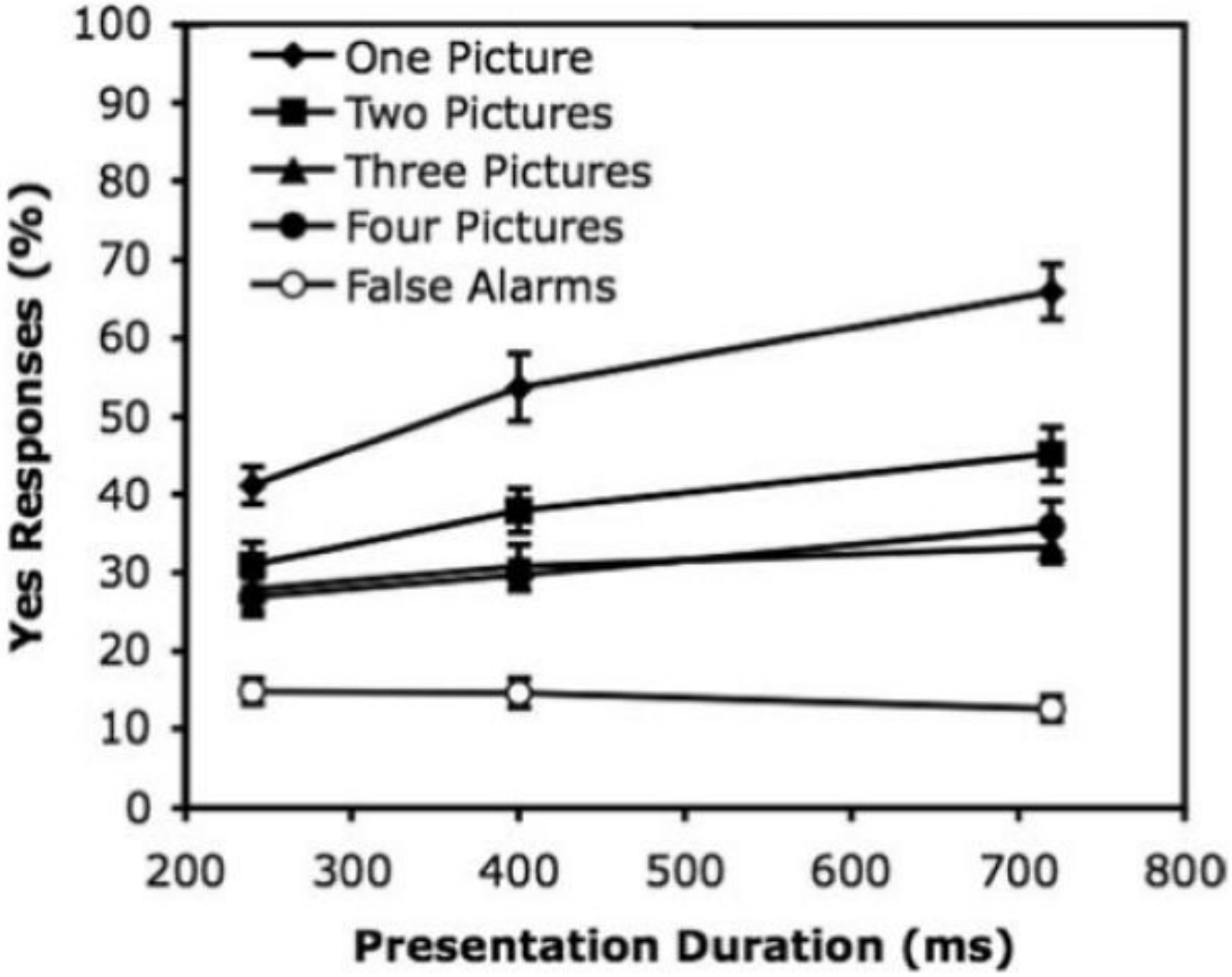

Figure 3.
Experiment 2: Percentage of correct *yes* responses in the recognition memory test of pictures at each presentation duration, separately for pictures from frames with 1, 2, 3, or 4 pictures. The percentage of false *yes* responses to distractor pictures is also shown for each presentation duration. Error bars represent standard error.

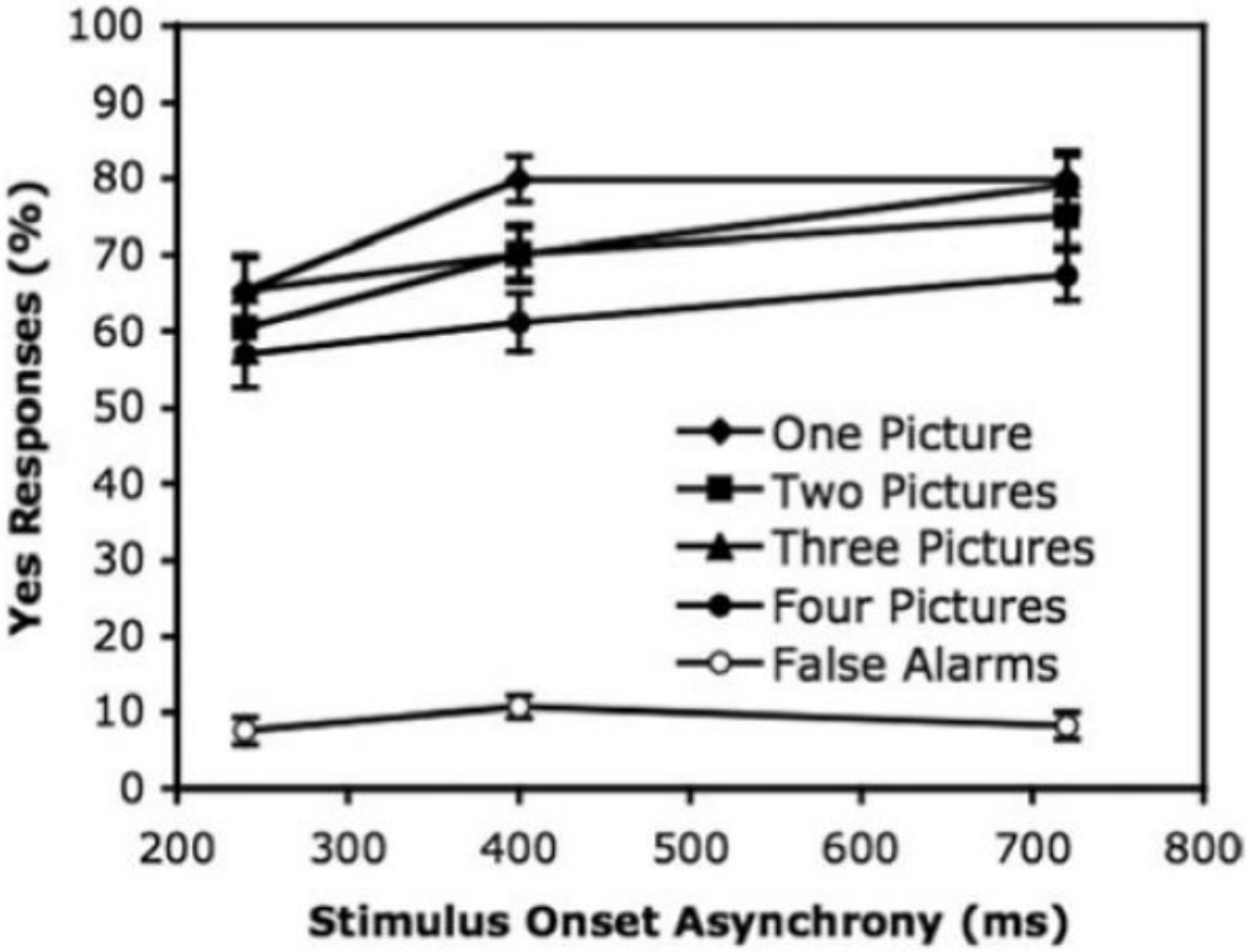

Figure 4.
Experiment 3: Percentage of correct *yes* responses to targets at each stimulus onset asynchrony (SOA; a duration of 160 ms plus a blank interstimulus interval), separately for target frames with 1, 2, 3, or 4 pictures. The percentage of false *yes* responses on target-absent sequences is also shown for each SOA. Error bars represent standard error.

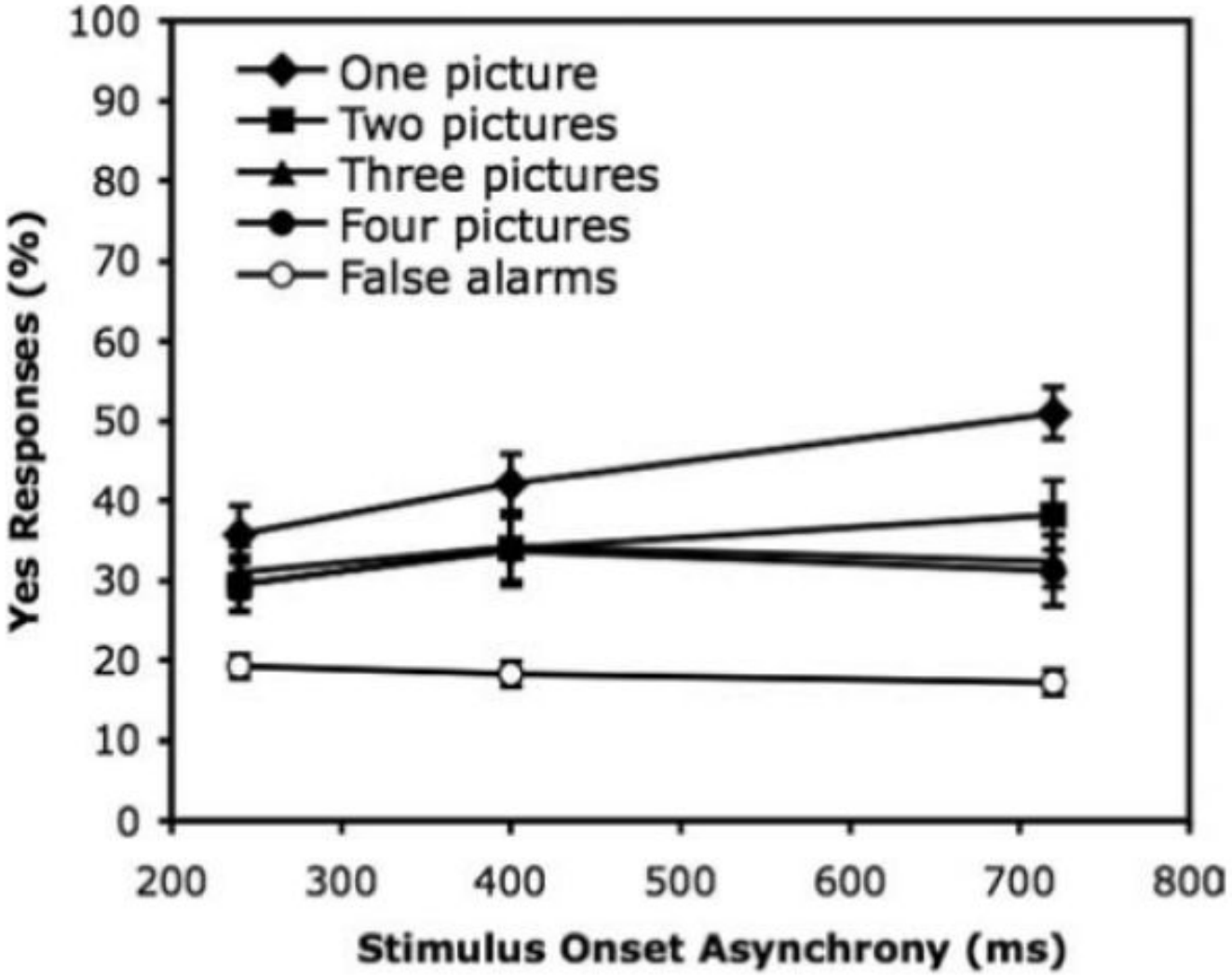

Figure 5.
Experiment 4: Percentage of correct *yes* responses in the recognition memory test of pictures at each stimulus onset asynchrony (SOA; a duration of 160 ms plus a blank interstimulus interval), separately for pictures from frames with 1, 2, 3, or 4 pictures. The percentage of false *yes* responses to distractor pictures is also shown for each SOA. Error bars represent standard error.