The efficiency of encoding: limits of information transfer into memory

Gesche M. Huebner · Karl R. Gegenfurtner

Abstract We investigated how visual information is best presented to maximize the number of remembered items in a fixed time unit. In a memory task for images depicting real-world objects, we varied the number of images shown simultaneously, the presentation time, and the interstimulus interval (ISI). The viewing phase was followed by a two-alternative forced choice recognition task. We converted the percentage of correct answers into a capacity estimate scaled to a fixed time unit of 1 s to allow for comparisons across conditions. Our results showed that (1) presenting one image very briefly was always more efficient than simultaneously showing multiple images for longer periods; (2) for single images, the maximum encoding rate was fairly constant over a wide range of conditions, at 1.4 objects per second; (3) when testing was done a week later, memory capacity was the same for all conditions, irrespective of the presentation time and ISI at the initial viewing; (4) highly similar distractors led to worse performance than random distractors; and (5) showing an image twice for 100 ms was associated with worse performance than showing the image once, but for 200 ms.

Keywords Efficiency · Memory · Natural objects · Presentation time · Interstimulus interval · Capacity

Human visual long-term memory (LTM) has an impressive capacity for images: Thousands of images can be stored after seeing them for several seconds (Shepard, 1967; Standing, 1973; Standing, Conezio, & Haber, 1970). For example, Standing et al. (1970) showed participants 2,560 slides for 10 s each and found that even after several days, recognition memory was still above 90%. However, in these experiments it took a long time simply to present the sequence of images to the observers. We were wondering just how quickly these images could be presented to achieve a maximally efficient encoding of the items into memory. By *efficient*, we mean that a certain level of performance is reached while presenting a number of images in the least amount of time.

There is a good deal of evidence that humans might need much less than 10 s to successfully parse an image and encode it into memory. Human observers are astonishingly fast and accurate in detecting prespecified targets in a stream of images. Thorpe, Fize, and Marlot (1996) briefly presented images to an observer, who had to detect the presence of an animal in the scene. Minimum reaction times were quite fast, and averaged brain potentials showed differences for target and nontarget images after only 150 ms. Given that the first evoked responses took place after about 70 ms, the authors inferred that it only takes about 80 ms of cortical processing to make this decision.

These two remarkable abilities—a vast memory capacity and ultrafast recognition—work on very different time scales. Here, we ask how they can be combined in order to most efficiently create memory representations. We aimed at finding the optimal parameters for presenting images to maximize the encoding of information into memory in a given time unit. Potential parameters that can be varied are the number of images presented simultaneously, the length of the presentation time, and the duration of the interval between stimuli (ISI). We will first summarize previous results on how these parameters affect rapid image processing and image consolidation, and then present the aims of our experiments.

G. M. Huebner (✉) · K. R. Gegenfurtner
Psychologie und Sportwissenschaft,
Abteilung Allgemeine Psychologie,
Fachbereich 06, Justus-Liebig-Universität Gießen,
Otto-Behaghel-Strasse 10F,
35394 Giessen, Germany
e-mail: gesche.huebner@gmail.com

Like many previous authors, we define successful comprehension and memorization as a correct answer in a two-alternative forced choice (2-AFC) recognition task in which the target and distractor are conceptually and visually dissimilar (except in Exp. 3). This means that a rather coarse representation of the object shown was sufficient for a correct answer, and our conclusions mainly hold true for such a distinction. Ideally, one would want to identify optimal presentation conditions for all types of distinctions (e.g., also for those between highly similar images). However, as long as there is no agreed-upon measure for the similarity of natural images, this aim is impossible to achieve. Therefore, we only touch upon this issue briefly, in one experiment (Exp. 3).

Rapid image processing

Picture comprehension can be seen as a necessary precursor to picture memorization. In the following section, we present results from rapid serial visual presentation (RSVP) studies that have given an estimate of the minimum time needed for understanding an image. Potter (1976) showed that with single images shown in an RSVP setup, target pictures were detected at a very high level of performance when the duration of a single frame was only about 113 ms. In these studies, participants received a verbal or pictorial cue about the to-be-detected image prior to seeing a rapidly presented stream of images. The high level of successful target detection despite the fast rate of presentation indicates that participants understood the meaning of images rapidly. Intraub (1981) showed that target detection was also very high when cues were nonspecific, such as a categorical cue that specified the superordinate category of the target and negative cues that specified what the target was not. Participants were not able to simply rely upon a template of the prespecified target, but they apparently comprehended the individual images. Further support for this remarkable ability to comprehend pictures after brief exposure came from Thorpe et al. (1996). They showed that after a presentation time as short as 20 ms, about 70 ms of cortical processing was sufficient for accurate detection of categorical targets such as animals. These findings have been expanded to nonbiological categories such as means of transport (VanRullen & Thorpe, 2001) and scene gist (Rousselet, Joubert, & Fabre-Thorpe, 2005; Rousselet, Macé, & Fabre-Thorpe, 2003).

Observers can also successfully discriminate between target and nontarget trials when carrying out an unrelated attention-demanding task. This result shows that rapid visual categorization of natural scenes requires no, or only little, focal attention (Li, VanRullen, Koch, & Perona, 2002).

Grill-Spector and Kanwisher (2005) showed that humans are as accurate and fast in categorizing objects as at merely detecting the presence of objects. Moreover, performance increased similarly in both tasks with longer presentation times up to 100 ms, indicating that the tasks require the same amount of information and are supported by similar underlying neuronal processes.

Thus, target detection and object and scene categorization can be performed rapidly and with high accuracy. These results are important for our studies when taking picture comprehension as an essential preceding step to picture memorization. The results imply that a viewing time of 20 ms with a processing time of 80 ms, and thus in total about 100 ms, are needed for comprehension and detection of a target. The lower boundary of our presentation times is thus given by this estimate.

Target detection for multiple images

It is still a matter of debate whether visual stimuli are processed serially or in parallel. This issue is of importance for optimizing the memorization of images, the question addressed in the present study: If multiple items can be processed simultaneously in the same amount of time as single items, then optimal memorization performance may be established by presenting multiple target images at the same time. However, if simultaneously attending to and processing multiple targets only occurs at the cost of increased processing time, the optimal memorization scheme may only entail one singly presented image.

Rousselet, Fabre-Thorpe, and Thorpe (2002) presented either one or two photographs of natural scenes for 20 ms, with the task to judge whether an animal was present. The speed of "go" responses was unaffected by whether one or two images had been presented, indicating a mode of parallel processing. Another animal detection study in which up to four pictures were presented at the same time confirmed the notion of parallel processing (Rousselet, Thorpe, & Fabre-Thorpe, 2004). Target detection in the periphery occurred with the same accuracy, irrespective of whether one or two images were presented (Li, VanRullen, Koch, & Perona, 2005).

VanRullen, Reddy, and Koch (2004) reached a different conclusion from their target detection study, in which up to 16 pictures were presented at the same time. When the stimuli were visible for 200 ms followed by a mask, accuracy dropped as more images were presented, reaching almost chance level when more than eight images were shown. With unrestricted presentation time, an increase in the number of pictures presented simultaneously was accompanied by an increase in response time of 40 ms per picture. This was taken as evidence for a serial processing mode. In the latest study addressing the issue of parallel cost-free or serial image processing, Potter and

Fox (2009) examined in detail the effect of presenting multiple pictures in a target detection task. They presented participants with an eight-frame RSVP sequence in which each frame contained between zero and four images, with a total number of eight images in each sequence. Target detection was significantly better when the target was presented alone than when with any number of other images were presented simultaneously. The same pattern of results was observed when participants had to perform a recognition task instead of a target detection task. An increase in the number of simultaneously presented images was associated with a decrease in performance. This result argued against cost-free parallel processing.

To summarize, the issue of parallel or serial processing has not yet been resolved. Until now, there is no conclusive proof for either parallel (with or without costs) or serial processing of visual images. We investigate this question in our experiments by varying the number of images presented concurrently. If two images were remembered as well as one image, given an equal presentation time, this would indicate cost-free parallel processing and would make simultaneous presentation of several images more efficient.

Consolidation of an image

As reviewed above, target detection and gist extraction occur rapidly and with high accuracy. But how stable is the apparent understanding of the images? Is it sufficient for successful memorization, or is additional time needed for encoding and consolidation?

Potter (1976) tested not only target detection but also recognition memory, using an identical image presentation design. After seeing the image sequence, observers performed a yes–no recognition task on the 16 presented images and 16 distractors. Image recognition was poor, with only 11% of correct responses at an image presentation rate of 113 ms, as compared to 64% correct target detection responses. Recognition memory increased with slower presentation rates but was drastically below target detection performance in all conditions. The superiority of target detection to recognition memory led to the conclusion that pictures were identified rapidly but then immediately forgotten when processing time was too short. Intraub (1981) likewise reported that significantly more pictures were detected than remembered.

The masking induced by the subsequent pictures was critical for the "memory loss": When images were presented equally briefly but now preceded and followed by a visual noise mask, recognition performance was as high as target detection when the presentation time was 120 ms (Potter, 1976). From this result and earlier ones (Potter & Levy, 1969), the conclusion was drawn that an

image is understood within 100 ms and is then immune to visual masking. At least another 300 ms of processing is necessary for a memory representation stable enough to survive conceptual masking, as induced by another meaningful picture. Longer presentation of the stimuli was associated with better recognition performance, as demonstrated in a study by Potter and Fox (2009) in which images were shown for 240, 400, and 720 ms. To reach asymptotic performance levels, up to 1 s of viewing or processing time might be needed (Potter, Staub, & O'Connor, 2004). Memorization of an image thus requires more time than target detection, and performance increases with longer viewing times, up to about 1 s.

Increasing the length of a blank interstimulus interval separating images also has a positive influence on later recognition memory (Tversky & Sherman, 1975; Weaver, 1974). Gegenfurtner and Sperling (1993) showed that in a partial-report procedure, in which a probed row of letters had to be reported, performance leveled off at a mask delay of 100 ms. Vogel, Woodman, and Luck (2006) reported that 50 ms per item were sufficient for consolidation. Even longer delays did not provide an additional benefit.

However, in the studies of Gegenfurtner and Sperling (1993) and Vogel et al. (2006), memory was tested after each trial. In typical RSVP tasks, similar to our approach here, many more items have to be remembered over a longer time period; thus, more time for consolidation, and consequently a longer separation between images, might be needed.

Potter and Fox (2009) also examined the role of the length of the ISI. Each frame was visible for 160 ms and was followed by a blank ISI of 80, 240, or 560 ms. Performance increased with a longer ISI, albeit not as much as when the presentation time of the to-be-remembered images was increased in duration.

Thus, longer presentation times and longer ISIs are associated with better performance in memory tasks, with possibly a larger contribution of increasing presentation time. Both factors are therefore potential candidates for influencing the efficiency of encoding. Increasing their length will lead to better performance, but might have the reverse effect on efficiency. Thus, we need to find out where the gain in performance outweighs the longer time needed for presentation. We therefore varied both presentation time and ISI during presentation of the images in the present experiments in order to measure their impact on performance as percentage correct and on efficiency—that is, performance for a given time unit.

How detailed is the representation?

Upon viewing an image, participants might remember a variety of features, ranging from the pure gist to specific

visual details. As mentioned earlier, we mainly aim at identifying the most efficient presentation mode for memorization of images in terms of discrimination from random distractors. For this, a very basic understanding and remembering of the image would be sufficient. It could be the case that different presentation parameters would be needed to achieve maximally efficient performance for other types of memory queries—for example, remembering the orientation of an object.

A longer viewing time might allow extraction of more details. Loftus, Nelson, and Kallman (1983) showed that information acquisition varies over the time course of image viewing: Holistic information is extracted during the first fixation, and specific features are extracted later. More recently, Melcher (2001, 2006) reported evidence for a buildup of visual memory over longer presentation times. Hollingworth and Henderson (2002) showed that change recognition performance improved with the total duration of time spent fixating on a target prior to the occurrence of a change.

Brady, Konkle, Alvarez, and Oliva (2008) showed participants thousands of images for 3 s each and then tested memory for a subset of the images. They concluded that visual LTM has a massive capacity and is capable of storing many visual details. When the distractor picture in the recognition task was only a "state change" of the original target picture—in other words, when the distractor depicted the target only slightly altered in appearance—performance was significantly lower, albeit still very high. Assuming that the extraction of details continues, one might expect that these differences would be compensated for with a longer presentation time of the images used for a state discrimination task.

We examined the accumulation of information and the details extracted in two experiments: In Experiment 5, we used state distractors, as had been done by Brady et al. (2008), and compared memory performance for these images with performance for random distractors using a range of presentation conditions. In Experiment 8, we tested accumulation of information by presenting an image once for a progressively longer presentation time or presenting it twice for half the time on each presentation.

Aims of our study

Our aim was to find the most efficient way of transferring information about images into memory. We define *efficiency* as the amount of information transferred into memory in a fixed time unit. If, for example, one image is remembered in a trial of 250-ms duration and 1.5 images are remembered in a trial of 500-ms duration, then by our definition, the first presentation mode would be more efficient.

First, in Experiments 1A and 1B, we varied the number of items presented at a time. If cost-free parallel processing occurs, showing more items at the same time should be more efficient. Alternatively, there could be an advantage for presenting only one item at a time, as suggested by Potter and Fox (2009). Second, in Experiment 2, we systematically varied presentation time and ISI because those variables affect performance in a visual memory task and are therefore inherently related to the efficiency of encoding.

As a third point, we investigated the nature of the memory representations in Experiment 3 by using distractor images that were conceptually and visually similar to the targets in the recognition task.

We also examined the stability of memory representations. Shortening presentation time and ISI might be associated with a less stable memory representation that is sufficient for immediate testing but not for recognition after prolonged time periods. In Experiment 4, we therefore retested a subset of participants after a delay of a week. In Experiment 5, we inserted a delay of a couple of minutes between image viewing and testing and examined the effects on memory performance. As a last point, in Experiment 6 we tested whether repeated exposure to an image (e.g., twice for 100 ms each time) was superior to only one exposure, with the same total presentation time (e.g., 200 ms).

General method

In all experiments, we ensured that the lighting conditions were the same for all participants. The experiments were carried out in accordance with the rules of the local ethics committee.

Participants

All participants were students of Justus Liebig University Giessen. Both male and female students participated, but the data were not analyzed separately for gender. The age range was between 19 and 30 years. Participants received €8/h for participation, with the actual payout depending on how long the experiment lasted. All participants had self-reported normal or corrected-to-normal vision.

Equipment

The experiment was written in MATLAB using the Psychophysics Toolbox extensions (Brainard, 1997; Pelli, 1997). Images were presented on an Iiyama VisionMaster 513 (MA203DT) 21-in. CRT screen with a monitor resolution of 1,280×960 pixel and a refresh rate of 100 Hz. A chinrest to stabilize the head position of the participants was placed at a distance of 47 cm from the

screen, resulting in a visual field of 48.2°×36.9° of visual angle.

Stimuli

We used the images provided by Brady et al. (2008). These images depict everyday items such as remote controls, dollar bills, and food items. We only used images that stemmed from distinct categories, except where otherwise noted. Images were resized to fit into a 4.8° of visual angle square bounding box. Depending on the specific conditions of the experiment, one, two, four, or eight images were shown concurrently. Images were arranged on an invisible circle with a radius of approximately 13° of visual angle. When only one image was shown, it was presented in the center of the screen. When two images were shown, they appeared at the 3 and 9 o'clock locations; in the four-images condition, they were at the 3, 6, 9, and 12 o'clock locations; and in the eight-images condition, they appeared evenly divided on the invisible circle starting at 3 o'clock. Participants were told that within one block of trials—that is, for trials with the same number of images—the pictures would always be shown at the same locations. When possible, the stimuli were counterbalanced across conditions.

Procedure

The experiment consisted of two phases: a learning phase and a testing phase. In the learning phase, participants viewed a sequence of images presented on the screen. Details about presentation time (PT), ISI, and number of images are given in the Method sections of the individual experiments. Participants were instructed to remember the images in preparation for a 2-AFC recognition task. The learning phase was followed by the test phase. In the test phase, two images, the target and a distractor, were shown side-by-side on the screen with unlimited viewing time. Participants indicated which image they believed they had previously seen: To select the image on the left side, participants pressed the left arrow, and to select the one on the right side, they pressed the right arrow. The target object was randomly chosen to be positioned on the left or the right side. After the participants had made a choice, the images disappeared. No feedback was given. The next image pair appeared after 1,000 ms. Figure 1 shows an example of the experimental procedure.

The role of concurrently presented images

In Experiment 1A, we investigated the effect of the number of images shown concurrently. We presented one, two, four, or eight images simultaneously. The ISIs were set to 500
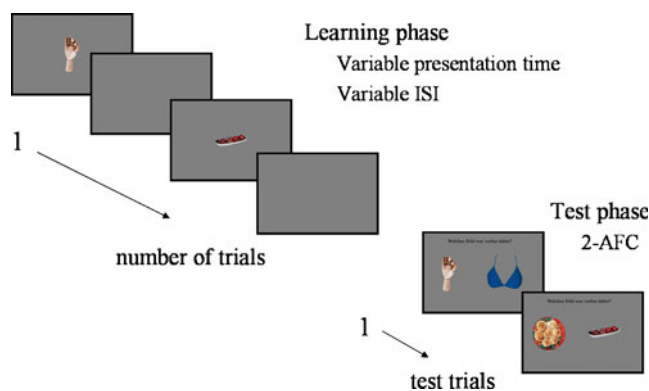

Fig. 1 Example of the experimental procedure. The learning phase was followed by the test phase, consisting of a 2-AFC recognition task

and 2,000 ms. The basic PTs were 250 or 500 ms per image, meaning that these values were multiplied by the number of images presented concurrently to arrive at the PT for a given condition.

By comparing conditions in which the PT and the ISI were the same but the number of objects varied, we examined the effect of number of images presented concurrently. In addition, this design allowed us to test whether variations in the PTs matter. For this analysis, we compared performance for different PTs with the same ISI. As a final point, we tested whether the length of the ISI mattered. For this examination, we compared performance for conditions that only varied in the ISI (500 vs. 2,000 ms).

In Experiment 1B, we shortened the presentation durations further for the multiple-images condition, to test whether this would lead to worse memory performance. For the one-image condition, we used longer and shorter trial durations to define the range in which variations mattered.

Experiment 1A

Method

Every participant performed four blocks of trials. The number of images shown simultaneously was held constant in every condition for a participant, whereas the PT and the ISI varied. The order of the conditions was randomized across participants. When one image was shown, the PT was 250 or 500 ms. When two images were shown, PTs were 500 or 1,000 ms. When four images were shown, PTs were 1,000 or 2,000 ms, and for eight images, they were 2,000 or 4,000 ms. There were always 5 participants for each number of images, except when only one image was shown ($n = 4$). Every condition consisted of 100 trials, except for the eight-images conditions, which consisted of 50 trials.

Results

Memory performance and capacity First, we report memory performance (as percentage correct) and memory capacity for the various conditions. Memory performance defined as the percentage of correct answers in the memory task for the different conditions is summarized in Table 1.

For one image, performance increased with longer trials. Performance was best for the combination of a long PT and a long ISI (500 and 2,000 ms, respectively) with $M = 92.75\%$. In this and all subsequently reported repeated measures ANOVAs, degrees of freedom were adjusted according to Greenhouse–Geisser. A repeated measures ANOVA with the factors PT and ISI showed significant main effects of PT, $F(1, 3) = 24.92$, $p = .015$, and ISI, $F(1, 3) = 94.27$, $p = .002$. The interaction was not significant. Thus, a longer presentation time and a longer ISI led to higher performance.

There were no significant main effects or interactions in any of the other conditions. Varying the PT and the ISI did not have a significant effect when multiple images were presented.

The main aim of our experiments was to examine the efficiency of encoding. To facilitate a comparison of results and to estimate the efficiency, we converted performance as percentage correct into a measure of capacity in terms of stored items. This measure also takes the number of simultaneously presented items into account. First, to correct for guessing, we calculated p, the corrected proportion correct: $p = (x - g)/(1 - g)$, where x is the raw proportion correct and g is the guessing probability of .5

(Busey & Loftus, 1994). The resulting p was then multiplied by the number of objects presented (e.g., one, two, four, or eight). Since our aim was to find the most efficient mode of presentation, we took the trial length into account. Thus, we calculated the capacity for a time unit of 1,000 ms. To do this, we divided the reference time unit of 1,000 ms by the trial duration of a condition. This gave us the factor with which we then multiplied the capacity estimate in the respective condition. For example, when the total trial length was 750 ms, the capacity estimate was multiplied by 4/3, whereas when it was 1,500 ms, the multiplier was 2/3. The results are shown in Fig. 2.

Depicted is the mean capacity per time unit and the standard error of the mean (*SEM*). The first bar stands for the combination of the respective shorter PT with an ISI of 500 ms, the second bar stands for the combination of the respective shorter PT and an ISI of 2,000 ms. The third and fourth bars represent the combination of the longer PTs and the two ISIs.

Using a long ISI was not profitable, as shown by the low values of the second and fourth bars. Performance per time unit was better when only one or two images were presented. Because for two images the tested variations in PT and ISI did not matter for performance, it is necessary to investigate whether and how shortening those times affected performance.

Memory performance: The role of the number of images To address whether images were processed in parallel without any costs—that is, without a drop in performance—we

Table 1 Performance (defined as percentage correct) for the various combinations of the number of concurrently presented images, presentation time, and interstimulus interval (ISI), Experiment 1A

| No. of Images	Presentation Time (ms)	ISI (ms)	Trial Duration (ms)	Performance (% correct)	Standard Deviation
1	250	500	750	76.00	8.98
1	250	2,000	2,250	85.50	9.26
1	500	500	1,000	83.50	7.59
1	500	2,000	2,500	92.75	6.34
2	500	500	1,000	66.00	7.34
2	500	2,000	2,500	71.80	12.44
2	1,000	500	1,500	73.20	14.53
2	1,000	2,000	3,000	79.60	8.50
4	1,000	500	1,500	60.00	3.94
4	1,000	2,000	3,000	62.20	8.08
4	2,000	500	2,500	66.80	10.35
4	2,000	2,000	4,000	72.00	9.03
8	2,000	500	2,500	50.40	4.04
8	2,000	2,000	4,000	57.20	4.32
8	4,000	500	4,500	56.80	7.10
8	4,000	2,000	6,000	56.40	4.04
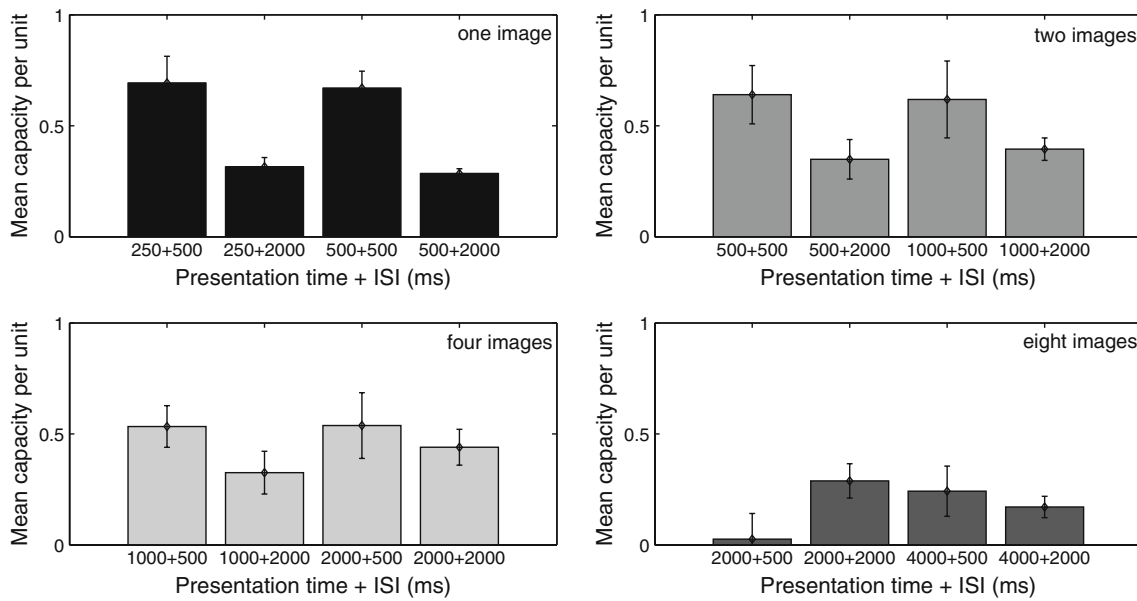
Fig. 2 Mean capacity per unit and *SEM* for the various combinations of presentation time and interstimulus interval (ISI), Experiment 1A. The labels in the upper-right corners specify the number of concurrently presented images

compared those conditions in which the number of objects varied but viewing time and ISI were identical. Performance was compared using a Welch test. Table 2 shows which conditions were compared and the resulting t and p values. The mean difference was calculated by subtracting performance for the participants who had seen more pictures from that of the participants who had seen fewer pictures presented concurrently. All differences were positive, showing that performance was always higher in the condition with fewer items presented. All differences were significant except for the conditions with PT 1,000 ms and ISI 500 ms with two or four objects presented.

From these results, we conclude that processing of multiple images does not occur in a cost-free parallel fashion: When the PT and ISI were identical, performance was lower when more items were presented.

With respect to the efficiency of encoding, the results showed that presenting one or two objects for a short time

was superior to presenting more images concurrently for a longer time.

Experiment 1B

In the second part of Experiment 1, we were mainly concerned with addressing the question of the most efficient encoding. Thus, we shortened the trial durations for multiple images presented concurrently. To estimate the range in which variations in PT and ISI mattered for one image presented, we used both longer and shorter trial durations.

Method

When one image was shown at a time, the total trial durations were 200 ms (PT 100 + ISI 100), 500 ms (PT 250 + ISI 250),

Table 2 Conditions and results of the Welch test used to estimate the effect of the number of images presented concurrently, Experiment 1A

No. of images (A)	No. of images (B)	Difference A−B (% correct)	Presentation time (ms)	ISI (ms)	t	p
1	2	17.50	500	500	3.485	.012*
1	2	20.95	500	2,000	3.272	.016*
2	4	13.20	1,000	500	1.960	.112
2	4	17.40	1,000	2,000	3.318	.011*
4	8	16.40	2,000	500	3.300	.020*
4	8	14.80	2,000	2,000	3.306	.017*

1,000 ms (PT 250 + ISI 750), and 1,250 ms (PT 750 + ISI 500). Four students took part in those four conditions. An additional 4 students participated in the other three conditions, in which either two images were shown (PT 250 + ISI 250; PT 500 + ISI 250) or four (PT 500 + ISI 500).

Results

Memory performance Table 3 summarizes performance in terms of the percentage correct for the various conditions.

For one object, a repeated measures ANOVA showed a significant main effect of PT, $F(3, 9) = 11.32$, $p = .035$. Pairwise comparisons with Bonferroni adjustment showed only one significant difference: Performance for a PT of 750 ms with an ISI of 500 ms was significantly higher than for a PT and ISI of 100 ms each. When only the ISI was different, with an identical PT of 250 ms, performance did not differ significantly, indicating that variations of an ISI between 250 and 750 ms did not have an effect.

For two objects with a PT of 250 ms and an ISI of 250 ms, mean performance was $M = 59.75\%$ ($SD = 4.86$). For the combination of a PT of 500 ms and an ISI 250 ms, mean performance was $M = 57.75\%$ ($SD = 2.75$). When four objects were shown, performance was $M = 55.5\%$ ($SD = 4.44$). The differences between the three conditions were not significant. The average performance in the two-images conditions in Experiment 1B was significantly lower than the average performance for two images in Experiment 1A, as a Welch test showed, $t(7) = 4.22$, $p = .007$, meaning that performance declined with shorter presentation times. The same conclusion held true for the four-images condition, $t(7) = 2.66$, $p = .033$.

Memory capacity Again, we looked at performance for a specific time unit. Figure 3 shows the capacity estimates for the various conditions. For one image (black bars), the unit capacity was highest when PT and ISI added up to 200 ms, with $M = 1.9$ objects remembered for the time unit of 1,000 ms, followed by $M = 1.01$ objects per time unit for a trial duration of 500 ms, and only $M = 0.62$ and $M = 0.63$ objects per time unit for the comparatively long blocks.

Thus, memorizing images is most efficient when individual images are presented for short durations. A repeated measures ANOVA showed a significant main effect of condition, $F(3, 9) = 21.45$, $p = .009$. Pairwise comparisons with Bonferroni adjustment showed that the capacity estimate in the shortest trial (200 ms) was significantly higher than in all other trials (for the 1,000-ms condition, $p = .030$; for 1,250 ms, $p = .048$; for 500 ms, $p = .033$).

For two images, the most objects per time unit ($M = 0.78$) were stored in the condition with the shortest trial duration (500 ms). For the other condition in which two images were shown simultaneously, only $M = 0.41$ images were memorized per time unit. For four images, the average was even lower, with $M = 0.36$ images remembered.

Discussion: Experiments 1A and 1B

Experiments 1A and 1B showed that variations in PT and ISI only had an effect when one image was presented. Although performance (as percentage correct) decreased with shorter trial durations, this effect was negligible in terms of the efficiency of encoding. The maximal efficiency was reached with one image presented for 200 ms (PT + ISI). To reach the same unit capacity for two images using a viewing time of 200 ms, it would be necessary for performance in that condition to be identical to performance in the 500-ms condition, which is highly unlikely; thus, we focused in subsequent experiments on presenting one image at a time. Presenting more images at one time while keeping PT and ISI constant was associated with worse performance, indicating that no cost-free parallel processing of images occurred.

The roles of PT and ISI

Experiment 1 showed that it was most efficient to present just one image for a very short time. We therefore focused on presenting one image at a time and conducted a stringent

Table 3 Performance (defined as percentage correct) for the various combinations of the number of concurrently presented images, presentation time, and interstimulus interval (ISI), Experiment 1B

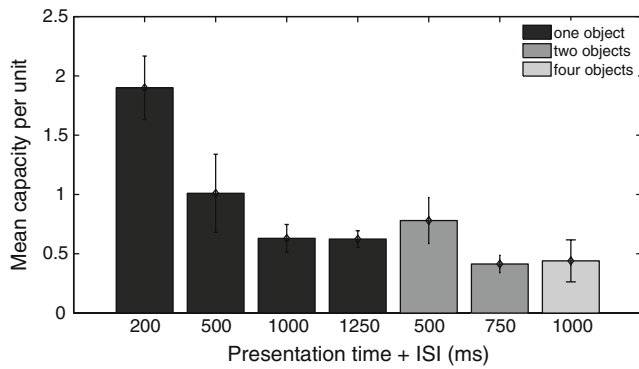| No. of images	Presentation time (ms)	ISI (ms)	Trial duration (ms)	Performance (% correct)	Standard deviation
1	750	500	1,250	89.00	8.79
1	250	750	1,000	81.50	11.71
1	250	250	500	75.25	16.46
1	100	100	200	69.00	5.35
2	250	250	500	59.75	4.86
2	500	250	750	57.75	2.75
4	500	500	1,000	55.50	4.44

Fig. 3 Mean capacity per unit and *SEM*, Experiment 1B. The values on the *x*-axis are the sums of the respective presentation times and interstimulus intervals (ISIs)

test to determine the PT and ISI at which we would find the highest capacity estimate per time unit.

Experiment 2

Method

We used PTs of 50, 100, 200, and 300 ms and ISIs of 0, 50, 100, and 200 ms, resulting in 16 conditions. A total of 20 students participated in all 16 conditions. Each condition consisted of 50 trials, presented in a blocked manner.

The PT was not shortened further, because exploratory work had shown that performance broke down to chance level when an image was only visible for 25 ms (with ISIs of both 25 and 50 ms).

Results

Memory performance Table 4 summarizes the conditions used in this experiment and gives the results. Performance is reported both as percentage correct and as capacity per time unit.

Inspection of Table 4 shows that performance tended to increase with longer trial durations. Performance was significantly above chance in all conditions, as shown by a one-sample *t*test. To disentangle the contributions of PT and ISI, we conducted a repeated measures ANOVA with these two factors. Both factors had a significant effect on performance, as did the interaction [PT, $F(3, 57) = 78.19$, $p < .000$; ISI, $F(3, 57) = 35.34$, $p < .000$; PT × ISI interaction, $F(9, 171) = 3.78$, $p = .003$]. The estimated marginal means for the four PTs were $M = 57.83\%$ (50 ms), $M = 62.53\%$ (100 ms), $M = 71.38\%$ (200 ms), and $M = 77.3\%$ (300 ms). Pairwise comparisons with Bonferroni adjustment showed that all differences were significant (all *p*s < .005). The estimated marginal means for the various ISIs were $M = 62.15\%$ (0 ms), $M = 65.55\%$ (50 ms), $M = 68.33\%$ (100 ms), and $M = 73.0\%$ (200 ms). Pairwise comparisons with Bonferroni adjustment showed that the differences between an ISI of 0 ms and all other ISIs were significant (all *p*s < .018). Thus, the shortest ISI was associated with worse performance. Performance for the longest ISI (200 ms) was significantly better than for all other ISIs (all *p*s < .004). The difference between ISIs of 100 and 200 ms was not significant.

This experiment showed a significant interaction between PT and ISI. For a PT of 50 ms, only the difference between an

Table 4 Performance (defined as percentage correct) and unit capacity for the various combinations of presentation time and interstimulus interval (ISI), Experiment 2

Presentation time (ms)	ISI (ms)	Trial duration (ms)	Performance		Capacity	
			% Correct	SD	Per Unit	SD
50	0	50	55.70	5.92	2.28	2.37
100	0	100	56.00	6.39	1.20	1.28
200	0	200	63.10	9.28	1.31	0.93
300	0	300	73.80	11.22	1.59	0.75
50	50	100	55.70	7.20	1.14	1.44
100	50	150	59.60	6.82	1.28	.091
200	50	250	69.40	7.63	1.55	0.61
300	50	350	77.50	12.45	1.57	0.71
50	100	150	58.30	7.41	1.11	0.99
100	100	200	62.40	11.58	1.24	1.16
200	100	300	74.60	9.25	1.64	0.62
300	100	400	78.00	10.99	1.40	0.55
50	200	250	61.60	7.33	0.93	0.58
100	200	300	72.10	9.05	1.47	0.60
200	200	400	78.40	9.19	1.42	0.46
300	200	500	79.90	11.42	1.20	0.46

ISI of 0 and 200 ms was significant ($p = .032$). For a PT of 100 ms, performance at an ISI of 200 ms was significantly higher than performance at any of the other ISIs (all $ps < .006$). For a PT of 200 ms, all differences were significant (all $ps < .030$), except for the comparison of ISIs of 100 and 200 ms. For a PT of 300 ms, only one difference in ISIs was significant (50 vs. 200 ms). Figure 4 shows performance as a function of PT and ISI.

The results indicate that for both the shortest and longest PTs, variations in the ISI were of lesser importance than in the other two conditions. For a PT of 50 ms, this could reflect baseline performance, meaning that such a short PT simply does not allow for higher performance, independent of the ISI. For a PT of 300 ms, the limited impact of the ISI might indicate saturation: The maximum performance was already reached during image presentation, meaning a blank interval did not bring an additional benefit.

Memory capacity We calculated performance in terms of the items stored per time unit of 1,000 ms. As Fig. 5 shows,
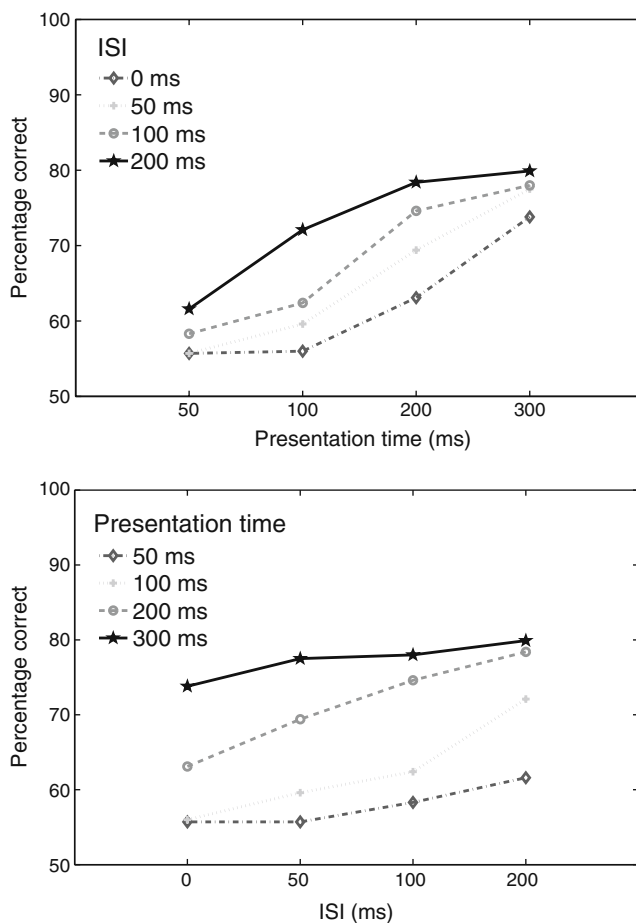

the mean unit capacity was rather similar across conditions, with the exception of a peak when total trial duration was only 50 ms (PT 50 + ISI 0). However, this cannot be considered as the most efficient way, because some people are unable to memorize anything in this condition; the variation in performance is very large.

We conducted a repeated measures ANOVA across all conditions. The effect of trial duration was only slightly above significance, $F(15, 285) = 2.39$, $p = .063$. In a repeated measures ANOVA with two factors, PT and ISI, neither of the factors reached significance (both $ps > .153$). The interaction was significant, $F(9, 171) = 7.09$, $p = .032$, but was limited in its scope: Significant effects only emerged for a PT of 300 ms; also, an ISI of 200 ms was related to a significantly lower unit performance than with any of the other ISIs (all $ps < .016$).

Analysis of viewing order We analyzed performance depending on the order in which the images had been viewed. For this analysis, in each condition we compared performance for the first, the middle (25th), and the last (50th) image seen in the sequence. We used Cochran's test for the comparison. Since 20 students had participated in each condition, the comparisons were made based on 20 data points for each condition. Only one condition (PT 250 + ISI 100) showed a significant effect, $p = .050$, in terms of a primacy effect: The first item that had been seen was remembered with a greater accuracy. In all other conditions, performance was the same for all trial numbers under consideration.

We were also interested in seeing whether performance on subsequent trials was related. Does consolidation of one image at trial n have a negative impact on the memorization of the subsequently presented image $n + 1$? One could imagine that consolidation might need more time than was provided by the ISI: When one image was consolidated successfully (as indicated by a correct answer in the testing session), consolidation would then have continued to take place while the next image was already shown, making it likely that the second image would not be remembered. In the whole sequence of trials, the combination of two subsequent correct trials should then be less likely to occur than the combination of an incorrect and a correct trial, or vice versa. We counted the numbers of correct–incorrect, correct–correct, incorrect–correct, and incorrect–incorrect pairs, and then compared these numbers against the expected numbers for each pairing that could be derived from the probabilities of a correct and an incorrect answer in the respective condition. Using a chi-square test, we compared the expected and the observed frequencies. We failed to find any significant effects, meaning that there was no clustering of certain outcomes for two successive trials.

Fig. 4 In the top graph, the four lines stand for the different interstimulus interval (ISIs), whereas the *x*-axis shows the different presentation times. In the bottom graph, these two factors are reversed. The *y*-axes show the percentages of correct answers, Experiment 2

Fig. 5 Mean capacity per unit and *SEM* sorted in order of increasing trial lengths, Experiment 2. The details of the presentation time (PT) and the ISI are as follows (PT + ISI, all in milliseconds): 50 + 0, 100 + 0, 50 + 50, 100 + 50, 50 + 100, 100 + 100, 200 + 0, 200 + 50, 50 + 200, 300 + 0, 200 + 100, 100 + 200, 300 + 50, 300 + 100, 200 + 200, and 300 + 200
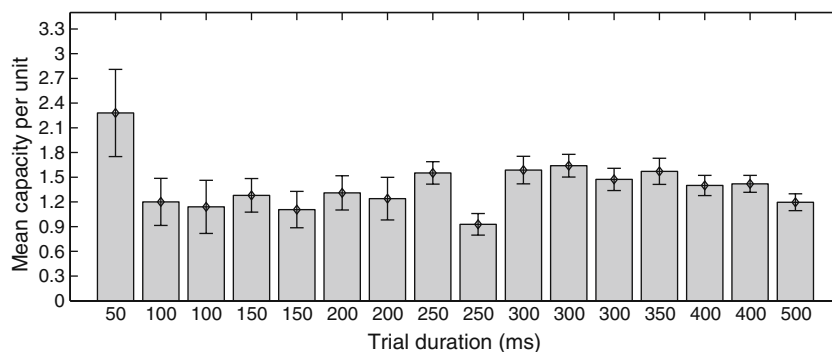
Discussion

In this experiment, we examined how variations in the PT and the ISI affect performance and the efficiency of encoding. We showed that increases in PT and ISI are associated with better performance, defined as percentage correct. Efficiency was hardly affected by these variations. We concluded that the rate of encoding into memory is fixed at about 1.4 objects per second. When the trial duration is extremely short (50 ms), very high capacity values can be reached. However, considering that only a subset of the participants were able to perform above chance, in general this is not the most efficient presentation mode, which is reflected by the fact that performance with a 50-ms trial duration was not significantly higher than in the other conditions. Clearly, for a subset of participants, very rapid presentation is the most efficient way of presentation. However, in any scenario that is not tailored to an individual but to a group, such as an advertisement on TV, this presentation mode is not more efficient than the others.

Additionally, we found a significant interaction effect (PT × ISI), with variations in the ISI mattering less at the longest and shortest viewing durations of the stimuli, possibly indicating baseline and saturation performance.

We did not find evidence for a reliable differential effect of item position in the learning phase. Neither the images viewed early nor late in the sequence were remembered better.

The role of distractor similarity

As reviewed in the introduction, there is evidence supporting the notion that participants are able to remember many visual details (Brady et al., 2008; Castelhano & Henderson, 2005; Hollingworth, 2004; Hollingworth & Henderson, 2002; Vogt & Magnussen, 2007). Up to this point in our experiments, the distractor image in the recognition task was always conceptually dissimilar to the target image. This means that participants could have relied upon storing solely a brief verbal label without remembering any visual details. It might well be that more time is needed to extract more details, and that consequently the most efficient

presentation conditions may vary. To investigate this issue, we used a particular set of images in the recognition task: the image pairs during testing—that is, the target and distractor images showed the same object, only in a different state (see Brady et al., 2008). For example, an image could depict a whole cookie, and the matching "state distractor" would depict the same cookie in a partially eaten state. Using these images, we aimed to test the extent to which participants stored a coarse, verbal label capturing the meaning of an image, or whether they stored visual details as well.

Experiment 3

Method

A group of 8 students participated in one part of this experiment, performing in four blocks of trials. We used a 2 × 2 design in which PT and distractor type were varied: The PT was set to either 100 or 200 ms, with a constant ISI of 50 ms, and the distractor in the memory task was either a random or a state distractor. A *random* distractor meant that the target image was paired with a categorically different distractor. A *state* distractor was a distractor image showing the same object as the target image, but in a different state. We used a blocked design, meaning that within one block, all images were shown for the same amount of time, and all trials in the recognition phase required the same type of discrimination (i.e., between a target and a random distractor or between a target and a state distractor).

Another 8 students took part in a further four conditions: (1) PT 100 + ISI 100, (2) PT 300 + ISI 100, (3) PT 400 + ISI 200, and (4) PT 800 + ISI 200. An interleaved design was used here: The image sequence consisted of 25 state distractors and 25 random distractors. This was done in order to avoid the use of specific strategies and to prevent potential frustration for the participants if discrimination between a target and a state distractor was too difficult. We used longer PTs to test the hypothesis that extracting the high level of detail necessary for performing a state

discrimination would need more time but could be done as well as a coarser discrimination, given a sufficiently long viewing duration.

Results

Memory performance The mean percentages correct for state distractors were $M = 53.5\%$ ($SD = 4.38$) for the trial duration of 150 ms and $M = 57.0\%$ ($SD = 10.37$) for the trial duration of 250 ms.

In both conditions, performance as percentage correct was slightly above the chance level (one-sample t test: for PT = 100 ms, $t(5) = 2.263$, $p = .058$; for PT = 200 ms, $t(5) = 1.910$, $p = .098$). Performance between the two conditions was not significantly different. When the distractor image was a random one, the respective mean percentages correct were $M = 61.0\%$ ($SD = 5.13$) and $M = 72.5\%$ ($SD = 5.43$). A paired-samples t test showed that performance was significantly higher for the random distractors, $t(7) = 3.071$, $p = .018$, and $t(7) = 5.305$, $p = .001$, respectively, for the 100- and 200-ms PTs. Thus, performance was much worse when using highly similar distractor images.

In the second set of conditions, performance was also lower when discrimination between a target and a state distractor was required than when the distractor image was conceptually dissimilar to the target image. Figure 6 shows performance defined as percentage correct for state distractors (light gray bars) and for random distractors (dark gray bars).

For the shortest trial duration (200 ms), the mean percentage correct for state images was $M = 58.0\%$ ($SD = 13.86$), and for random images it was $M = 68.5\%$ ($SD = 13.08$). For the trial duration of 400 ms, the mean percentage correct for the state images was $M = 75.0\%$ ($SD = 9.01$), as compared to $M = 81.5\%$ ($SD = 13.68$) for the random images. When images were shown for a PT of 400 ms with an IS of 200 ms, the respective mean percentages correct were $M = 82.5\%$ ($SD = 11.89$, state

images) and $M = 89.0\%$ ($SD = 11.06$, random images). For the longest trial duration of 1,000 ms, performance was almost unchanged for the state distractors ($M = 83.5\%$, $SD = 11.79$) but increased again for the random distractors ($M = 94.0\%$, $SD = 5.24$).

We analyzed performance using a repeated measures ANOVA with the factors Presentation Duration and Distractor Type. The main effect of presentation duration was highly significant, $F(3, 21) = 20.88$, $p < .000$. The main effect of distractor type was also significant, $F(1, 7) = 11.76$, $p = .011$; for state distractors, performance was lower than for random distractors. The interaction was not significant. Pairwise comparisons with Bonferroni adjustment showed that performance for the shortest duration was significantly lower than in all other conditions, all $ps < .018$. Performance for a PT of 300 ms with an ISI of 100 ms was significantly lower than performance for a PT of 400 ms with an ISI of 200 ms ($p = .023$) but not different from performance for the longest-duration condition (PT 800 + ISI 200). Performance in the two longest-duration conditions did not differ significantly.

We tested performance as percentage correct against the chance level of 50% using a t test. In all conditions performance was significantly above chance, except for state distractors with a PT of 100 ms and an ISI of 100 ms, $t(7) = 1.63$, $p = .149$.

When converting performance as percentage correct into the capacity per time unit, a repeated measures ANOVA showed once again the main effect of distractor type, $F(1, 7) = 6.64$, $p = .037$. The interaction between viewing duration and distractor type was not significant, though, showing that higher efficiency was not reached with longer viewing durations for state distractors. As a tendency, capacity per time unit declined with longer viewing durations; for random distractors, it was highest for the shortest condition of 200 ms ($M = 1.8$), whereas for the state distractors it was highest for the 400-ms condition ($M = 1.25$).

Fig. 6 Mean percentage correct and *SEM* for the different combinations of presentation time and interstimulus interval (ISI) for random (dark gray) and state (light gray) distractors, Experiment 3
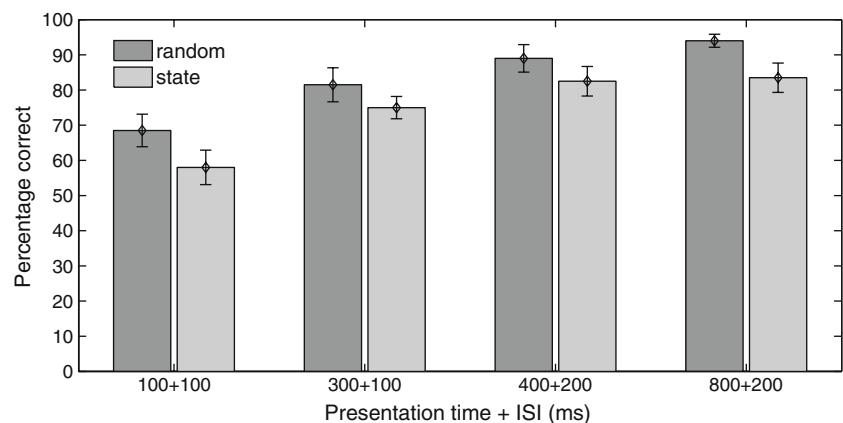
Discussion

This experiment showed that general performance and unit capacity were significantly lower for state distractors. Thus, some level of detail is either lost or not initially encoded during the time course of a trial sequence. To reach above-chance performance in a recognition task where the target and distractor are highly similar, the PT and ISI need to be between 100 and 300 ms. Even for the longest trial duration, performance was lower for state distractors than for random distractors and did not increase beyond the performance level of the second-longest condition, suggesting an asymptote of performance and a general disadvantage for state distractors. With respect to the memory representation, the results suggest that the level of detail extracted increases with longer PTs but reaches an asymptote that is lower than when coarse distinctions have to be made.

The stability of the memory representation

There is one issue in our studies that has not yet been addressed—namely, the stability of the representation: When the trial length was shorter, the testing also occurred sooner, meaning that the information had to be stored for a shorter time period. Designing an experiment that could tackle this issue—ensuring comparable storage conditions, independent of the parameters PT and ISI—is challenging. Introducing a blank delay of such duration as to achieve the same total durations is problematic, since it would allow for active rehearsing of the stimuli for different amounts of time, thus creating another confound. Likewise, an in-between task varying in length to reach the same total durations might again differentially impact performance. We opted for a retest of a subset of participants one week after the first testing session. We assumed that the seconds or minutes in which the image sequences had differed would not matter anymore when a second test was performed after such a long delay. As a second way to address this question, we introduced a delay of several minutes between image viewing and testing for a subset of conditions and compared performance from the immediate and delayed testing conditions.

Experiment 4: Retest after a week

Method

The 8 participants from the Experiment 2 (reported above) were recruited for a follow-up test session that was separated from the original session by about one week.

Upon arrival, the participants were informed that they would be tested again on the images they had seen in the first session. The original target images (e.g., those that made up the image sequence) were paired with novel images. We did not use the previously shown distractor images because those had also been seen and might have left a memory trace. Performance during the first testing of the participants selected for the retest had not been significantly different from that of the other participants.

Results and discussion

Table 5 shows initial performance (T1) and retest performance (T2) and the mean difference (T2–T1) for the various conditions.

We analyzed performance using a repeated measures ANOVA with the factors Testing Session and Condition. There was a highly significant main effect of testing session, $F(1, 7) = 20.44$, $p = .003$. The estimated marginal mean for performance was higher in the first testing session, with $M = 66.36\%$ versus $M = 58.56\%$ in the second testing session. The effect of condition was also highly significant, $F(1, 7) = 7.10$, $p = .002$. Performance for the trial durations of 400 ms (PT 300 + ISI 100 or PT 200 + ISI 200) and 500 ms (PT 300 + ISI 200) was significantly higher than performance for trial durations of 100 ms (PT 100 + ISI 0 or PT 50 + ISI 50). Performance for the combination of a PT of 200 ms and an ISI of 200 ms was significantly higher than performance for a PT of 50 ms with an ISI of 0 ms. Most importantly, the interaction was also significant, $F(1, 7) = 3.15$, $p = .026$. We plotted the performance for the first and second testing sessions in Fig. 7; significant differences are marked with an asterisk.

Pairwise comparisons for the different conditions at the two testing points showed significant differences in performance (as percentage correct) between the following conditions, mostly those with longer trial durations; the first number indicates the PT, the second number the ISI: 200 + 50, 200 + 100, 200 + 200, 300 + 0, 300 + 50, 300 + 100, or 300 + 200 ms. When considering only the second testing session, there were no significant differences in performance between all 16 conditions. The results indicate that performance in the longer conditions decreased to a baseline level, similar to that observed for the short trial durations in initial testing. This means that the amount of information that can be transferred to a more permanent store is the same for all conditions. Importantly, performance did not break down over the delay, as Fig. 8 shows. Even though there was little room for a decrease in performance for the shortest condition, where the mean was only slightly above chance, no forgetting with respect to the first testing seemed to have occurred.

Calculating the capacity per trial averaged across conditions showed that $M = 0.17$ objects were retained in

Table 5 Performance (defined as percentage correct) in the different combinations of presentation time (PT) and interstimulus interval (ISI) for first and second testings (T1 and T2) of the participants in Experiment 4, and the mean difference

PT (ms)	ISI (ms)	Trial Duration (ms)	Performance T1 (% correct)	Performance T2 (% correct)	Mean Difference T1–T2
50	0	50	55.25	56.00	−0.75
100	0	100	56.50	58.00	−1.50
200	0	200	62.25	58.25	4.00
300	0	300	73.25	60.50	12.75
50	50	100	53.75	53.25	0.50
100	50	150	61.00	52.50	7.75
200	50	250	72.50	59.50	13.00
300	50	350	77.25	60.75	16.50
50	100	150	57.25	56.75	.50
100	100	200	62.75	56.75	6.00
200	100	300	74.50	57.50	17.00
300	100	400	74.50	59.75	14.75
50	200	250	59.25	57.50	1.75
100	200	300	70.25	62.25	8.00
200	200	400	75.00	63.50	11.50
300	200	500	76.50	64.25	12.25

memory. To emphasize that performance was still substantially above zero images remembered, we plotted the capacity for a trial sequence in Fig. 8, meaning that we multiplied the respective capacity per trial by 50, taking into account that 50 images were presented per condition. Thus, on average, it can be inferred that $M = 8.65$ objects were still available in memory, per condition.

Experiment 5: Delayed versus immediate testing

Method

We tested 10 participants in an additional experiment in which we used six different timing conditions. The conditions were PT 100 + ISI 0, PT 200 + ISI 0, PT 100 + ISI 100, PT 200 + ISI 100, PT 200 + ISI 200, and PT 300 + ISI 200. All participants ran each condition twice: once with a delay of about 2 min between image viewing and the testing phase, and once with testing occurring immediately after the image viewing. During the delay period, the participants read a one-page-long story. They were told that at the end of the experiment they would have to rate the stories with respect to their valence. Half of the participants first did the six conditions without a delay followed by six with the story reading; the other half participated in the opposite order. The orders of the different PTs and ISIs were randomized across participants but were kept the same for each participant in the delayed and the immediate testing. After the experiment, participants were asked to rate the stories with respect to their

Fig. 7 Mean percentage correct and *SEM* for the first (broken gray line) and second (black line) testing sessions, in order of trial duration
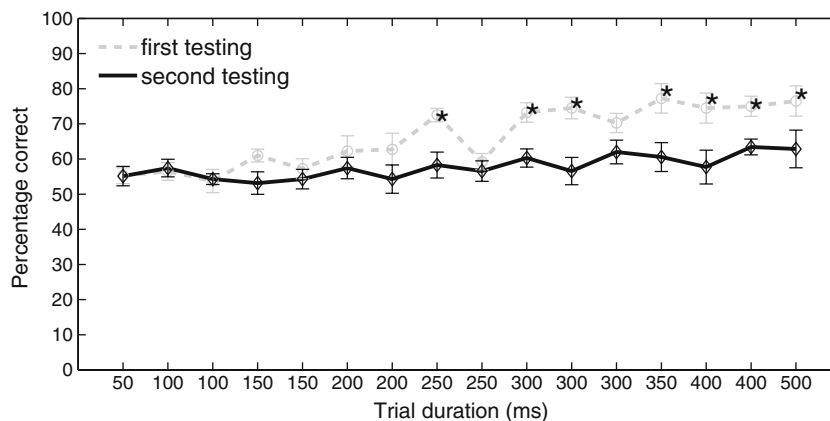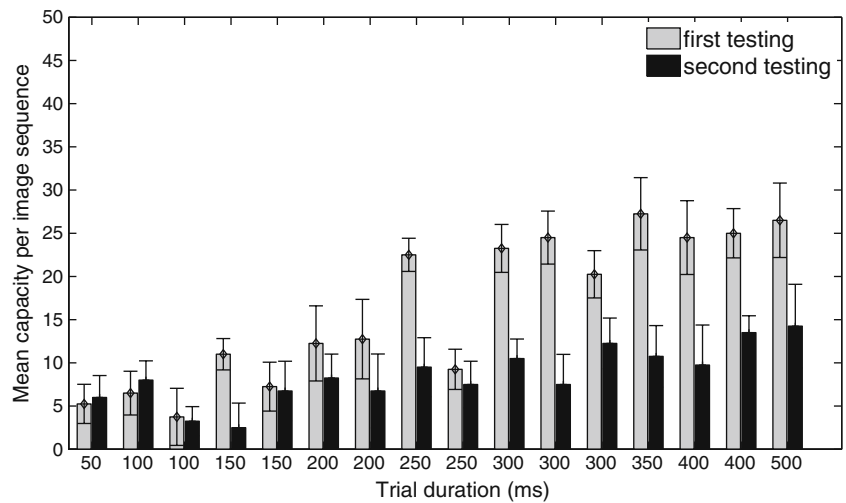
Fig. 8 The mean number of images remembered per image sequence for the first (gray bars) and the second (black bars) testing sessions

valence in an oral format. They were also told that the stories served the purpose of introducing a delay between image viewing and testing and that the valence ratings would not be used further.

Results

We analyzed the percentage of correct answers for every condition. Inspection of Fig. 9 shows that performance in the immediate (light gray) and delayed (dark gray) testing conditions was very similar.

A repeated measures ANOVA showed that there was no main effect of delay condition, $F(1, 9) = 0.32$, $p = .585$. The main effect of trial length was significant, $F(5, 45) = 40.63$, $p < .000$. Pairwise comparisons supported the earlier results: Longer PTs and ISIs were associated with better performance. Performance for the two longest conditions (estimated marginal means: 76.7% for PT 200 + ISI

200 ms, 79.50% for PT 300 + ISI 200 ms) did not differ from each other, but both were significantly higher than performance in all other conditions, all $ps < .028$. Performance in the shortest condition (PT 100 + ISI 0) was significantly lower than in the conditions with PT 200 + ISI 0 ms ($p = .003$) and with PT 200 + ISI 100 ms ($p > .000$). The last significant difference was between the condition with PT 100 + ISI 100 ms (estimated marginal mean 60.9%) and PT 200 + ISI 100 ms ($M = 69.7\%$), $p = .007$. The interaction between delay condition and trial length was not significant, $F(5, 45) = 0.54$, $p = .683$.

Discussion

The manipulation of introducing a delay between image viewing and testing had no impact on performance. This result indicates that a comparably stable memory trace is available in both immediate and delayed testing. We take this as evidence that we were measuring a LTM component, as discussed in the General Discussion.

Double versus single exposure

In all previous experiments, we kept the number of images that were seen in one sequence the same (except for Exp. 1). The duration of a sequence thus varied depending on PT and ISI. In this experiment, we set out to test performance for image sequences with equal total PTs, but with an unequal number of images shown, and accordingly different PTs and ISIs per image. Our previous results showed that the specific conditions were of little importance for unit capacity, but it could still be that an advantage exists for shorter trial durations: Within a given time unit, an image could be shown repetitively, which might be associated with higher performance than single showing.
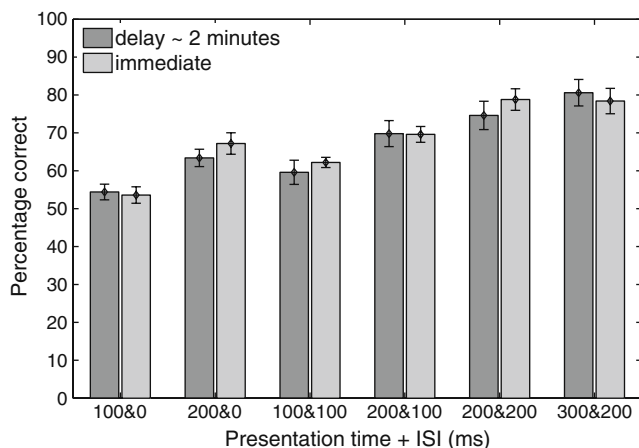

Fig. 9 Mean percentage correct and SEM for the immediate (light gray) and the delayed (dark gray) testing sessions for the six different conditions, in order of trial duration

Experiment 6

Method

A total of 8 students participated in all four conditions. Two of the conditions were always matched. In the first, *repeat-rapid*, condition, participants viewed 100 images for 100 ms, separated by an ISI of 100 ms. Note that, because there were only 50 distinct images, each image was shown twice. Testing was done on all 50 images. In the matched condition (*single-slow-1*), 50 images were shown only once for 200 ms, with an ISI of 200 ms. Thus, the overall duration of both sequences was the same, as was the total exposure time to each individual image, but the images were seen either twice or once for a doubled duration. The images shown and tested in each condition were counterbalanced across participants.

In a third condition (*single-rapid*), 100 different images were shown for 100 ms, with an ISI of 100 ms. Only 50 of the images were used in the testing session, to keep the session equal in length to the other conditions. The matched control condition (*single-slow-2*) consisted of the presentation of 50 images for 200 ms, with an ISI of 200 ms. Again, which images were used in which condition was counterbalanced across participants.

The durations of the PT and ISI were chosen to be in the previously found range of greatest efficiency.

Results

When each image was shown twice (PT 100+ ISI 100; repeat-rapid), the performance as percentage correct was $M = 78.0\%$ ($SD = 10.25$). In the corresponding control condition (single-slow-1), in which 50 images were shown only once but with doubled PT and ISI, the percentage correct was $M = 88.0\%$ ($SD = 6.41$). A paired-samples t test showed that performance in the single-exposure condition was significantly higher than in the repeat condition, $t(7) = -3.669$, $p = .008$. Thus, prolonged single exposure to an image led to better performance than repeated shorter exposures.

When 100 images were shown rapidly without repetition (single-rapid), the percentage correct was $M = 67.0\%$ ($SD = 9.44$). In the corresponding control condition (single-slow-2; 50 images, PT 200 + ISI 200), the performance as percentage correct was $M = 84.0\%$ ($SD = 8.41$). A paired-samples t test showed that performance in the condition with longer exposure per image was significantly higher than in the single-rapid condition, $t(7) = -4.106$, $p = .005$.

Comparison between the two conditions in which PT and ISI were identical (repeat-rapid vs. single-rapid) showed that performance was significantly higher when images

were shown twice, $t(7) = 2.453$, $p = .039$. Thus, there was a significant beneficial effect of repeated exposure.

Comparing the two matching conditions for showing images repeatedly or showing twice as many images (single-slow-1 vs. single-slow-2) allowed for estimating whether the images used were equally well remembered. These conditions were identical in terms of PT (200 ms) and ISI (200 ms), but different images were used. Performance here did not differ significantly, $t(7) = 1.497$, $p = .178$. This result indicates that the images were remembered equally well, allowing for comparisons across the conditions.

We converted percentages correct into capacity estimates. We multiplied the capacity per trial by the total number of images presented—that is, 50 in all conditions except the single-rapid condition. This served as a test for whether it was more efficient to show 100 images briefly or 50 images for double the time.

When 50 images were shown twice (repeat-rapid), $M = 28$ images ($SD = 10.25$) were stored on average. In the corresponding control condition (single-slow-1), $M = 38$ images ($SD = 6.41$) were remembered. Of the 100 images shown once (single-rapid), $M = 34$ images ($SD = 18.88$) were retained. In the control condition (single-slow-2), $M = 33$ images ($SD = 7.40$) were remembered.

The difference between the rapid-repeat condition and the corresponding control condition was significant, $t(7) = -3.669$, $p = .008$. Fewer items were remembered after brief double showings than after a single but longer presentation. When 50 or 100 images were shown in a sequence of the same total duration, the estimates did not differ significantly, $p = .835$.

Discussion

An image shown twice for 100 ms was remembered worse than an image shown once for 200 ms. Repeated presentation of the same image did increase performance, however, in comparison to single presentation for 100 ms. Apparently, more details can be extracted with longer *uninterrupted* viewing time. The result that performance increased for images viewed twice for a short time, as opposed to single short viewing, implies that some features extracted from first viewing are stored and that other information is taken up during second viewing: To explain the higher performance for images viewed twice, it seems reasonable to assume that different features are extracted during second viewing, with respect to first viewing.

The calculated capacity estimates indicated that if one wants to ensure that specific content is encoded, one should show the material once and opt for a longer trial duration. However, when only considering how much is encoded, the presentation mode does not matter; capacity estimations did not differ significantly between the condition with 100

images presented once for a PT and ISI of 100 ms and the one with 50 images presented with a PT and ISI of 200 ms each.

General discussion

We examined memory performance for rapidly presented images. Participants viewed a sequence of 50 or 100 images and then performed a 2-AFC recognition task. Our main aim was to find the most efficient way of presenting stimuli—that is, to maximize the amount of information stored in a given time unit. To investigate this issue, we varied the presentation time of the stimuli, the interstimulus interval separating the images, and the number of images presented simultaneously. We studied the effect of these variations on both memory performance in terms of percentage correct and memory capacity for a given time unit. We also addressed the issue of the stability of the resulting memory representations. As a final point, we investigated the nature of the memory representations.

The roles of PT and ISI for performance and efficiency

When presenting one item at a time, we found that shortening the PT and the ISI was generally associated with a decrease in performance defined as the percentage of correct answers. A trial length of 100 ms was necessary for above-chance performance, given that the distractor images were dissimilar to the target images. Our main aim was to address the efficiency of encoding—in other words, to find the optimal combination of PT and ISI to transfer the most information into memory in a fixed time period. When converting the data to this measure, the PT and ISI proved to be of little importance. Independent of presentation mode, 1.4 objects were remembered in 1,000 ms, meaning that one object was encoded in approximately 750 ms. This estimate might seem long, considering that Potter and colleagues (Potter, 1976; Potter & Levy, 1969) had estimated 300–400 ms for creating a lasting memory trace. However, our image sequence consisted of 50 images. Potter, Staub, Rado, and O'Connor (2002) showed that performance declined when more than 20 images were presented in one sequence. Considering that we used even more images, it might not be surprising that we found overall lower performance and a lower rate of encoding. Presenting an image twice for half the time was associated with poorer performance than showing it once, indicating that uninterrupted viewing and the accompanying uninterrupted consolidation was of greater benefit for memorization. The presentation mode thus mattered when specific content had to be memorized.

The role of the number of images

We also varied the number of images presented concurrently. We found that performance was best when presenting only one image at a time. When increasing the number of images while keeping the PT constant, performance declined. This argues against cost-free parallel processing. This was true even when the number of images was within the capacity limitations of visual short-term memory (VSTM) commonly reported as encompassing four items (e.g., Cowan, 2000; Luck & Vogel, 1997). In general, the question of whether serial or parallel processing occurs in target detection has not been resolved (Rousselet et al., 2002, 2004; but see Potter & Fox, 2009; VanRullen et al., 2004). Our results support the notion that there is no cost-free parallel processing.

Stability of the memory representation

A retest of a subset of participants after one week showed that for initially longer trial durations, performance dropped to the level of performance reached in trials with shorter durations in the first testing. In these short conditions, no further decline in performance was observed. This means that, irrespective of the specific way of presenting information, about the same amount of information is transferred into a more permanent store. Consequently, when seeking the most efficient way of presenting information for more permanent storage, shorter times are superior to longer durations. Introducing a delay of several minutes between image viewing and subsequent testing did not affect performance.

How detailed is the representation?

Earlier work had shown that human observers can extract the category and the gist of a scene very quickly (e.g., Potter, 1976; Rousselet et al., 2005; Thorpe et al., 1996). Ample research on visual memory showed that the memory trace contains many visual details and consists of more than a broad conceptual description (Brady et al., 2008; Castelhano & Henderson, 2005; Hollingworth, 2004; Hollingworth & Henderson, 2002; Vogt & Magnussen, 2007).

Our results show that general performance was lower when the distractor images in the recognition task were highly similar to the target images than when they depicted distinctively different objects. The PT of the stimuli needed to be longer to achieve above-chance performance for these state distractors than for random distractors, indicating that more time is needed to extract details beyond the overall meaning. This result is supported by the finding that a single exposure for a given duration is associated with better performance than double exposure for half the duration. To extract more details, more uninterrupted viewing time is needed. However, the details extracted

from first viewing are apparently used in selecting which information to extract during second viewing.

The finding that more time is associated with greater extraction of details also makes it necessary to limit our claims about the most efficient presentation conditions for discrimination between dissimilar images.

The type of memory

Visual memory is normally divided into three components: visual sensory memory, VSTM, and LTM. Consequently, one may ask what type of memory we measured. We can exclude visual sensory memory, a highly detailed yet extremely fleeting memory that is volatile to masking and to other interruptions, such as eye movements (Henderson & Hollingworth, 2003). In our experiments, subsequent images masked previous images, and participants certainly moved their eyes before testing started; thus, the observed memory performance was not attributable to visual sensory memory. With the same certainty, we can say that the memory capacity that was observed after one week of testing had to be ascribed to a long-term component.

But what about the capacity measured directly after image viewing? A large body of evidence suggests that the capacity of VSTM is limited to about four items (Alvarez & Cavanagh, 2004; Cowan, 2000; Irwin & Andrews, 1996; Luck & Vogel, 1997; Pashler, 1988; Todd & Marois, 2004). In our studies, these capacity limitations were exceeded in all conditions except for those in which performance was at chance. Even in the shortest conditions, at least six objects were remembered. To arrive at these estimations, the per-trial capacity was multiplied by the number of images presented, usually 50.

LTM has been shown to be involved after only a short image exposure. For example, in the context of scene viewing, Hollingworth and Henderson (2002) proposed that both STM and LTM contribute to the online representation of a scene. We have shown that LTM was involved in the memorization of a stimulus array that was visible for only 10 s (Huebner & Gegenfurtner, 2009). We assume that the memory performance measured in the present experiments was supported by an LTM component: Introducing a delay of several minutes between image viewing and testing did not affect performance relative to immediate performance. The delay ensured that we would test LTM performance, and the missing difference with respect to the immediate testing proved that also in the immediate testing condition we measured an LTM component. However, performance declined over a week, meaning that some forgetting occurred—it could therefore be proposed that two different components underlie the immediate and the one-week-delayed performance. Melcher (2001, 2006) suggested the existence of a medium-term component, a precursor of

LTM. It could thus be that the more immediate performance could be ascribed to the medium-term component and the performance after one week to LTM. More importantly, however, we did not measure visual STM, but rather a longer-term component.

Conclusion

The main aim of these experiments was to find the most efficient way of presenting visual information to reach maximum performance in terms of stored objects in a given time unit. We showed a general advantage of presenting only one object at a time, which argues against cost-free parallel processing. Apart from this effect, basically all combinations of PTs and ISIs led to the same capacity estimation of about 1.4 encoded objects per the time unit of 1,000 ms. Thus, there is a fixed rate of encoding into visual LTM that is not altered by varying the presentation mode. The stored representation lacks some detail, as indicated by the lower performance when testing was done on highly similar image pairs. Independent of presentation mode, the performance after a one-week delay was similar for all conditions: For short trials, performance did not decline further; for the longer trials, it decreased, but only to the point where it was comparable to the performance for shorter trials during first testing. Thus, the amount of information transferred into LTM is the same, irrespective of presentation mode.

References

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

Brady, T. F., Konkle, T., Alvarez, G. A., & Oliva, A. (2008). Visual long-term memory has a massive storage capacity for object details. *Proceedings of the National Academy of Sciences, 105*, 14325–14329.

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

Busey, T., & Loftus, G. (1994). Sensory and cognitive components of visual information acquisition. *Psychological Review, 101*, 446–469.

Castelhano, M., & Henderson, J. M. (2005). Incidental visual memory for objects in scenes. *Visual Cognition, 12*, 1017–1040.

Cowan, N. (2000). The magical number 4 in short-term memory: A reconsideration of mental storage capacity. *Behavioral and Brain Sciences, 24*, 87–185.

Davenport, J. L., & Potter, M. C. (2004). Scene consistency in object and background perception. *Psychological Science, 15*, 559–564.

Fei-Fei, L., Iyer, A., Koch, C., & Perona, P. (2007). What do we perceive in a glance of a real-world scene? *Journal of Vision, 7*(1), 10:1–29.

Gegenfurtner, K. R., & Sperling, G. (1993). Information transfer in iconic memory experiments. *Journal of Experimental Psychology: Human Perception and Performance, 19*, 845–866.

Grill-Spector, K., & Kanwisher, N. (2005). Visual recognition: As soon as you know it is there, you know what it is. *Psychological Science, 16*, 152–160.

Hollingworth, A. (2004). Constructing visual representations of natural scenes: The roles of short- and long-term visual memory. *Journal of Experimental Psychology: Human Perception and Performance, 30*, 519–537.

Hollingworth, A., & Henderson, J. M. (2002). Accurate visual memory for previously attended object in natural scenes. *Journal of Experimental Psychology: Human Perception and Performance, 28*, 113–136.

Huebner, G. M., & Gegenfurtner, K. R. (2009). Effects of viewing time, fixations, and viewing strategies on visual memory for briefly presented natural objects. *Quarterly Journal of Experimental Psychology, 63*, 1398–1413.

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

Irwin, D. E., & Andrews, R. (1996). Integration and accumulation of information across saccadic eye movements. In T. Inui & J. L. McClelland (Eds.), *Attention and performance XVI: Information integration in perception and communication* (pp. 125–155). Cambridge: MIT Press, Bradford Books.

Li, F.-F., VanRullen, R., Koch, C., & Perona, P. (2002). Natural scene categorization in the near absence of attention. *Proceedings of the National Academy of Sciences, 99*, 9596–9601.

Li, F.-F., VanRullen, R., Koch, C., & Perona, P. (2005). Why does natural scene categorization require little attention? Exploring attentional requirements for natural and synthetic stimuli. *Visual Cognition, 12*, 893–924.

Loftus, G. R., Nelson, W. W., & Kallman, H. J. (1983). Differential acquisition rates for different types of information from pictures. *Quarterly Journal of Experimental Psychology, 35A*, 187–198.

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

Melcher, D. (2001). Persistence of visual memory for scenes. *Nature, 412*, 401.

Melcher, D. (2006). Accumulation and persistence of memory for natural scenes. *Journal of Vision, 6*, 8–17.

Pashler, H. (1988). Familiarity and the detection of change in visual displays. *Perception & Psychophysics, 44*, 369–378.

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

Potter, M. C. (1976). Short-term conceptual memory for pictures. *Journal of Experimental Psychology: Human Learning and Memory, 2*, 509–522.

Potter, M. C., & Fox, L. F. (2009). Detecting and remembering simultaneous pictures in RSVP. *Journal of Experimental Psychology: Human Perception and Performance, 35*, 28–38.

Potter, M. C., & Levy, E. I. (1969). Recognition memory for a rapid sequence of pictures. *Journal of Experimental Psychology, 81*, 10–15.

Potter, M. C., Staub, A., & O'Connor, D. H. (2004). Pictorial and conceptual representation of glimpsed pictures. *Journal of Experimental Psychology: Human Perception and Performance, 30*, 478–489.

Potter, M. C., Staub, A., Rado, J., & O'Connor, D. H. (2002). Recognition memory for briefly-presented pictures: The time course of rapid forgetting. *Journal of Experimental Psychology: Human Perception and Performance, 28*, 1163–1175.

Rousselet, G., Fabre-Thorpe, M., & Thorpe, S. J. (2002). Parallel processing in high level categorization of natural images. *Nature Neuroscience, 5*, 629–630.

Rousselet, G. A., Joubert, O. R., & Fabre-Thorpe, M. (2005). How long to get to the "gist" of real-world natural scenes. *Visual Cognition, 12*, 852–77.

Rousselet, G. A., Macé, J.-M., & Fabre-Thorpe, M. (2003). Is it an animal? Is it a human face? Fast processing in upright and inverted natural scenes. *Journal of Vision, 3*, 440–455.

Rousselet, G. A., Thorpe, S. J., & Fabre-Thorpe, M. (2004). Processing of one, two or four natural scenes in humans: The limits of parallelism. *Vision Research, 44*, 877–894.

Shaffer, W. O., & Shiffrin, R. M. (1972). Rehearsal and storage of visual information. *Journal of Experimental Psychology, 92*, 292–296.

Shepard, R. N. (1967). Recognition memory for words, sentences, and pictures. *Journal of Verbal Learning and Verbal Behavior, 6*, 156–163.

Standing, L. (1973). Learning 10,000 pictures. *Quarterly Journal of Experimental Psychology, 25*, 207–222.

Standing, L., Conezio, J., & Haber, R. N. (1970). Perception and memory for pictures: Single trial learning of 2560 visual stimuli. *Psychonomic Science, 19*, 169–179.

Thorpe, S., Fize, D., & Marlot, C. (1996). Speed of processing in the human visual system. *Nature, 381*, 520–522.

Todd, J. J., & Marois, R. (2004). Capacity limit of visual short-term memory in human posterior parietal cortex. *Nature, 428*, 751–754.

Tversky, B., & Sherman, T. (1975). Picture memory improves with longer on time and off time. *Journal of Experimental Psychology: Human Learning and Memory, 1*, 114–118.

VanRullen, R., & Thorpe, S. J. (2001). Is it a bird? Is it a plane? Ultra-rapid visual categorisation of natural and artefactual objects. *Perception, 30*, 655–668.

VanRullen, R., Reddy, L., & Koch, C. (2004). Visual search and dual-tasks reveal two distinct attentional resources. *Journal of Cognitive Neuroscience, 16*, 4–14.

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

Vogt, S., & Magnussen, S. (2007). Long-term memory for 400 pictures on a common theme. *Experimental Psychology, 54*, 298–303.

Weaver, G. E. (1974). Effects of poststimulus study time on recognition of pictures. *Journal of Experimental Psychology, 103*, 799–801.