Memory for Repeated Images in Rapid-Serial-Visual-Presentation Streams of Thousands of Images

Evelina Thunell and Simon J. Thorpe
Centre de Recherche Cerveau et Cognition, Centre National de la Recherche Scientifique (CNRS), Université Toulouse III—Paul Sabatier

Abstract
Human observers readily detect targets in stimuli presented briefly and in rapid succession. Here, we show that even without predefined targets, humans can spot repetitions in streams of thousands of images. We presented sequences of natural images reoccurring a number of times interleaved with either one or two distractors, and we asked participants to detect the repetitions and to identify the repeated images after a delay that could last for minutes. Performance improved with the number of repeated-image presentations up to a ceiling around seven repetitions and was above chance even after only two to three presentations. The task was easiest for slow streams; performance dropped with increasing image-presentation rate but stabilized above 15 Hz and remained well above chance even at 120 Hz. To summarize, we reveal that the human brain has an impressive capacity to detect repetitions in rapid-serial-visual-presentation streams and to remember repeated images over a time course of minutes.

Keywords
rapid serial visual presentation, memory, learning, repetition detection, natural images, spike-timing-dependent plasticity

Received 5/28/18; Revision accepted 2/12/19

Throughout life, our brains adapt to our sensory environment. Attention and reinforcement are certainly important factors for brain plasticity and learning but are perhaps not strictly necessary. In fact, mere repetition of sensory input can be sufficient. In statistical learning, for example, regularities and repeating events or features can be learned implicitly without conscious awareness (Turk-Browne, 2012). For systems with very little knowledge about the world, such as the infant brain, repetition might be particularly important for deciding what to learn. In adults, fast learning of repeating auditory noise has been demonstrated (Agus, Thorpe, & Pressnitzer, 2010), and similar stimulation was later found to elicit “memory-evoked potentials” (Andrillon, Kouider, Agus, & Pressnitzer, 2015). These findings point to efficient repetition encoding in the auditory domain.

Here, we further explored the idea that repetition on short time scales is a key factor for learning. We set out to test the capacity of the visual system for detecting and remembering repetitions in streams of never-before-seen natural images, using a novel paradigm based on rapid serial visual presentation (RSVP; Potter, 1976; Potter & Levy, 1969). There is a vast body of research on RSVP using different types of items, such as words, letters, and natural images, which reveals a phenomenal capacity for processing briefly presented stimuli in rapid succession. Typically, the task is to attend to certain predefined target images, attend to a target category (such as “animals”), or remember all the images. The streams usually contain few items and last only seconds, making these tasks feasible.

Our RSVP paradigm differs in some key aspects from the classic ones. First, there were no predefined target
images or categories. Instead, participants were instructed to detect and remember any repeated images. These were chosen at random and appeared at random intervals, meaning that it was impossible to anticipate which image would become a target. Second, each stream lasted around 30 s to 80 s and contained hundreds or even thousands of images, making it impossible to memorize them all. Doing so would in any case not be useful, because the repeated items had to be identified among nonrepeated items from the same stream.

We demonstrated a remarkable sensitivity for repeated images across a wide range of parameters, with only two to three presentations being enough for the visual system to notice the repeat. This sensitivity does not depend solely on the total time of exposure but rather on the number of repetitions and on the number of distractors between repeats (as previously found in the auditory domain; Kaernbach, 2004; but see Martini & Maljkovic, 2009). Our findings raise questions about the underlying mechanisms capable of sustaining this ability. We propose that detection of repeating statistics, or even of exact repeating stimuli, is a fundamental property of our sensory systems that allows the brain to efficiently process sensory input.

Experiment 1

Method

Participants. Sixteen paid participants (mostly graduate and undergraduate students; 5 male) between the ages of 21 and 32 years (M = 24 years) took part in Experiment 1. All participants reported having normal or corrected-to-normal vision and gave written informed consent before the experiment. The procedures were in accordance with the Declaration of Helsinki and approved by the local ethics committee (Comité d’Évaluation Éthique de l’Inserm, or CEEI Protocol Number 2015-004).

Stimuli and task. Experiment 1 was aimed at testing the effect of the number of presentations of a repeated image on repetition encoding. Participants viewed RSVP streams of never-before-seen natural images presented at 15 Hz; each image was displayed for 58.3 ms, followed by an 8.3-ms gap (one screen refresh of background gray). The images were taken from a subset of 160,000 images acquired by sampling evenly from the 1,000 themed folders in the ImageNet database training set (http://images-net.org/), which contains over 1 million images. These images were cropped symmetrically from all sides so they were square in size (8.3 cm per side, thus spanning about 6.3° of visual angle) and were equalized in resolution to 150 × 150 pixels. We chose the gray background (RGB values = 114, 114, 114) to match the average gray level of a sample of 14,000 randomly chosen images from the set.

Each RSVP stream contained 18 repetition sequences, consisting of an image appearing between 2 and 10 times, interspersed with nonrepeating distractor images. In half of the streams, there was one distractor between each repeat, as shown in Figure 1, and in the other half, there were two distractors. Before each repetition sequence, there was an interval of between 2.5 s and 3.5 s (37–52 images) of nonrepeated images, making it impossible to foresee which image would be repeated next. Because of the random duration of these intervals, the total duration of the streams varied slightly. The one-distractor streams lasted 65.1 s to 70.3 s (977–1,054 images) with an average of 67.9 s (1,018 images). The two-distractor streams lasted between 71.3 s and 76.0 s (1,069–1,140 images) with an average of 73.8 s (1,107 images).

Participants were instructed to press “s” on the keyboard every time they detected a repetition sequence and to remember the repeated images for a subsequent memory task. They were encouraged to push the button as soon as they noticed there was a repetition of an earlier image, but they were told that this button press could be made after the repetition sequence had already finished or even after the image stream had finished.

After each stream, participants performed a four-alternative forced-choice memory task. Response frames of four images were presented, and participants were asked to indicate which image (the first, second, third, or fourth from the left) had been repeated in the preceding image stream. After choosing an image, participants rated how confident they were of their response on a scale from 0 (guessing) to 4 (completely sure). They were informed that one image in each response frame had been repeated and that the other three had also appeared in the stream, but only once. (A similar identification task is implemented in our freely available iPad game “Brainspotting”; https://itunes.apple.com/app/id1246763569.) The order of the targets in the memory task was randomized and thus did not match the order of their appearance in the RSVP stream. The two tasks were used to evaluate different aspects of repetition processing: the detection stage and memory. In addition, by linking the two tasks, we were able to assess the importance of conscious detection for memory. Above-chance performance on the detection task also ensured that participants were attending to the stream. No feedback was given during the experiment.

The experiment consisted of two sessions separated by 24 hr (± 1 hr); each session consisted of 20 blocks (RSVP stream + memory task). In each stream, there were two occurrences of each condition of between 2 and 10 target presentations (i.e., one to nine repetitions), appearing in a random order.
Fig. 1. Example stimuli and trial sequence in Experiment 1. Each rapid-serial-visual-presentation (RSVP) stream was presented at 15 Hz and contained approximately 1,000 images (left). At random time points, an image was repeated (red lines). In this example repetition sequence, there are five presentations of the target (the image of the plush toys). RSVP streams with one distractor between repetitions, as shown here, were interleaved with streams with two distractors. “RT = 0” indicates the onset of the second presentation (i.e., the first repetition) of the target. The RSVP stream was followed by a memory task (right), in which the repeated images had to be identified among sets of three nonrepeated images from the same stream. Each response frame stayed on screen until participants made their choice and was then followed by a prompt for a confidence rating.
Before starting the first session, participants practiced on four blocks: two blocks with 2-Hz streams containing five targets each, followed by two blocks with 15-Hz streams containing eight targets each. In the first block of each presentation rate, there was one distractor between repeats, and in the second, there were two distractors. The experimenter was in the room during the practice session and gave some verbal feedback (such as feedback on the memory task regarding whether answers were correct or incorrect, when possible). The purpose of the training was to familiarize participants with the task, to make sure they understood the instructions, and to encourage guessing in the memory task (participants often identified the right image even when they lacked confidence about their response). Before starting the experiment, participants were informed that the image streams would last longer than during the practice session.

Procedure. Each participant viewed approximately 42,000 images. Apart from the repeated images and the images in the response frames, no images were seen more than once by any participant. All images were equally likely to be used as a target, and the images selected as targets varied between participants. Half of the participants started with a one-distractor stream, and the other half started with a two-distractor stream.

Participants were seated approximately 75 cm from the screen in a dimly lit room (~0.3 cd/m²). The RSVP streams were presented in the middle of the screen. In the memory task, the four images were centered on the horizontal meridian of the screen, with approximately one third of the image width between images. Before each block, participants were reminded of the instructions, and the block number was indicated. Participants themselves launched each RSVP stream and were encouraged to take a short break before doing so, if needed. They also launched the memory task, which was preceded by a reminder of the task. Participants were instructed to look at the images in the streams but were free to fixate on any part of the image. Eye movements were not recorded.

Apparatus. The stimuli were presented on a 24-in. BenQ XL2411Z monitor (BenQ, Taoyuan, Taiwan) with a 120-Hz refresh rate (resolution: 1,920 × 1,080), controlled by a PC. We verified the reliability of the screen refresh rate and the stimulus-generating script using a photodiode connected to an oscilloscope.

Analysis. A repeated image in the RSVP stream was considered detected if there was one and only one button press at least 200 ms after the onset of the image’s second presentation (the first repetition) and not more than 2 s after the onset of its last presentation. The onset of the second presentation was defined as a reaction time (RT) of 0. We compared the detection performance with a chance level computed by performing the same analysis on scrambled data. Here, for each participant and block, the button presses were given random time stamps. When analyzing the memory-task data, we first considered all trials together. In a second analysis, we separated detected and missed targets. Naturally, in the latter case, the number of trials per data point varied between conditions and participants, ranging from zero to all.

The delay between the appearances of a particular target in the RSVP stream and in the memory task was defined as the number of intervening targets in the stream and in the memory task plus one. For example, for a target that was repeated last in the RSVP stream and prompted for first in the memory task, this value is 1. We split the data in half with respect to this delay and averaged within these two groups. The very longest and shortest delays were uncommon and occurred for only a few participants, and they were therefore not included. Short delays were 3 to 18 images, and long delays were 19 to 33 images. Note that the number of trials per delay condition varied between participants. Further, in this analysis, the data were pooled across conditions (number of presentations) in a random, non-balanced way.

We used MATLAB (The MathWorks, Natick, MA) and Psychtoolbox-3 (Brainard, 1997; Kleiner, Brainard, & Pelli, 2007) for stimulus generation and analysis. The 95% confidence intervals (CIs) for the participant means were computed using the MATLAB bootci function with 2,000 iterations and the bias-corrected-and-accelerated-percentile method (default). We used the freely available software JASP (JASP Team, 2018) to calculate Bayesian t tests.

Results

Repetition detection was easier the more times the image was presented, up to a ceiling at around seven (Fig. 2a). There is no easily defined false alarm rate in this paradigm. Instead, to compute a meaningful chance level for comparison, we performed the same analysis on scrambled data. Repetition-detection performance was well above this chance level in the one-distractor and two-distractor conditions, except when the image was presented only twice (see Fig. S1 in the Supplemental Material available online). The task was easier with one distractor than with two distractors (the 95% CIs for the difference between two distractors and one distractor fell below zero for all conditions; see Fig. S2a in the Supplemental Material). RTs remained rather stable across number of presentations in the one-distractor streams (Fig. 2a). In two-distractor streams,
RTs were longer overall and increased slightly with the number of presentations.

In the four-alternative forced-choice memory task that followed each RSVP stream, the repeated images had to be identified. Performance again improved with the number of presentations up to around seven. Interestingly, for the one-distractor streams, the CI did not span chance level even for images that had been shown only twice (Fig. 2b). Performance was higher in the one-distractor than in the two-distractor streams (for 6–10 presentations, the CIs for the difference between distractor types did not include zero; see Fig. S2b in the Supplemental Material). The improved identification performance for more presentations is not due to a speed/accuracy trade-off, because if anything, RTs decreased with the number of presentations (Fig. 2b). The confidence ratings mimic performance (Fig. 2b), indicating that participants correctly judged the relative difficulty levels of the different conditions.

Do repetitions have to be consciously perceived in order to be memorized? To answer this question, we analyzed the memory-task data separately for detected and missed targets. Interestingly, many of the repeated images that were not detected during the RSVP stream were still correctly identified in the memory task: The CIs for performance fell above chance for all conditions except for two presentations with two distractors (Fig. 3b). Confidence followed the same pattern (Fig. 3b), indicating that participants were likely aware of recognizing these images. For comparison, the corresponding results for the detected repetitions are shown in Figure 3a. For a comparison between Day 1 and Day 2, see Figure S6 in the Supplemental Material. Performance improved slightly from Day 1 to Day 2, with accompanying decreased RTs, in both the repetition-detection and memory tasks.

**Experiment 2**

**Method**

**Participants.** Twenty-four paid participants (10 male) between the ages of 21 and 32 years ($M = 26$ years) took part in Experiment 2; 7 of them had already participated.
in Experiment 1. The number of participants and trials in Experiment 2 were determined on the basis of a pilot study with image-presentation rates between 2 and 30 Hz conducted on 3 additional participants (including the authors).

**Stimuli and task.** Experiment 2 was aimed at testing the effect of the image-presentation rate (i.e., frequency) of the RSVP stream. The paradigm was the same as in Experiment 1, with the following exceptions. First, the number of presentations of each repeated image was fixed at five, and there were always two distractors between each presentation. Second, the presentation rate differed between blocks and could be either 4, 6, 8, 10, 12, 15, 17.1, 20, 24, 30, 40, 60, or 120 Hz (corresponding to interimage intervals of 250, 166.7, 125, 100, 83.3, 66.7, 58.5, 50, 41.7, 33.3, 25, 16.7, and 8.3 ms). Third, each stream contained 10 repetition sequences. Last, streams both with and without a gap between images (8.3 ms of background gray) were included. The RSVP streams lasted, on average, 40.5 s: from 31.1 s (3,738 images) for the 120-Hz streams to 62.5 s (250 images) for the 4-Hz streams.

The experiment consisted of two sessions that took place on two consecutive days at approximately the same time (within 45 min). Each session consisted of 26 blocks, so there were two streams of each of the 13 different presentation rates. For rates up to and including 60 Hz, there was a gap of 8.3 ms between images in one of the streams (gap conditions). Since

---

**Fig. 3.** Mean identification performance (top row) and confidence in identifications (bottom row), separately for (a) detected and (b) missed repetition targets in Experiment 1. Identification performance was measured as the percentage of correctly identified targets. The horizontal black line in the top row marks the 25% chance level. In all graphs, each data point represents the mean for 1 participant, and the curved lines represent the mean across participants. The shaded areas are 95% confidence intervals.
the screen-refresh rate was 120 Hz, both 120-Hz streams necessarily had no gaps (no-gap conditions).

As in Experiment 1, participants practiced the task before starting the experiment on the first day. There were two practice runs of 13 blocks increasing from 4 Hz to 120 Hz (only no-gap conditions). The practice streams contained two repeated images presented 10 times each in the first run and five times each in the second run. Before launching the experiment, we informed participants that the image streams would last longer than they had during practice and that the presentation rate would not necessarily increase monotonically across blocks.

**Procedure, apparatus, and analysis.** The procedure, apparatus, and data-analysis plan were the same as in Experiment 1, with the following differences. Each participant viewed approximately 51,000 images. For participants who took part in both experiments, we used nonoverlapping sets of images. Half of the participants viewed the gap streams first, and the other half viewed the no-gap streams first. The different image-presentation rates were presented in a random order. In this experiment, short trial delays were 3 to 10 images, and long trial delays were 11 to 17 images.

**Results**
The 8.3-ms gap of background gray, introduced to test the effect of image duration and of image onsets from a neutral background, seems to have had a negligible influence (see Fig. S7), and the gap and no-gap conditions were therefore pooled in all further analysis. Not surprisingly, participants detected the most repetitions in the slowest image streams; the difficulty increased with the presentation rate up to around 15 Hz (Fig. 4a). Above this rate, however, performance decreased less rapidly and remained well above chance up to 120 Hz (Fig. S8 in the Supplemental Material). RTs decreased with increasing image-presentation rate (Fig. 4a), as was to be expected, considering that the repetition sequences were shorter at higher presentation rates.

Performance in the memory task resembled that in the repetition-detection task—participants identified the most targets from the slowest streams but performed above chance up to 120 Hz (the CIs did not span chance level for any presentation rate; Fig. 4b). Identification RTs increased slightly with image presentation rate (Fig. 4b). As in Experiment 1, confidence ratings mimicked identification performance (Fig. 4b).

Identification performance for missed targets was lower overall than in Experiment 1, and the CIs fell above chance level only for the 4-, 6-, 8-, 12-, and 24-Hz conditions (Fig. 5). The confidence ratings reflect the performance difference between missed and detected targets (Fig. 5). For a comparison between Day 1 and Day 2, see Figure S12 in the Supplemental Material. The only visible difference across days was a slight RT decrease in the repetition-detection task.

After finishing the last session, participants were asked to describe their experience. Sixteen of the 24 participants spontaneously reported a "shining through" of the target at high image-presentation rates: Repeated images "persisted," "emerged in the background," "floated on top of the other images," or "stood out" and might be the reason for the above-chance performance in the faster streams. Thirteen participants said that they used specific strategies such as "fixating a spot" or "relaxing the eyes" to facilitate this special percept.

**Discussion**
Our results reveal an impressive capacity of the human brain for detecting repetitions embedded in RSVP streams containing hundreds or even thousands of never-before-seen natural images and for remembering the repeated images over a time course of minutes. While this study did not address the question of the neural correlates, it seems plausible to imagine some efficient repetition detectors involved in the early stages of memory encoding—for example, cells that are sensitive to certain features or objects and that will readily detect repetitions while ignoring nonrepeated items. One candidate mechanism for such learning is spike-timing-dependent plasticity (STDP). This simple and unsupervised Hebbian learning rule varies synaptic strength as a function of the relative timing of pre- and postsynaptic spikes and is capable of causing neurons to become sensitive to repeated input spike patterns (Bichler, Querlioz, Thorpe, Bourgoin, & Gamrat, 2012; Masquelier, 2018; Masquelier, Guyonneau, & Thorpe, 2008; Song, Miller, & Abbott, 2000). STDP has been suggested as a possible mediator of statistical learning (Goujon, Didierjean, & Thorpe, 2015), which is a more general form of repetition encoding.

Since repetition detection increased with the number of presentations, and every presentation of the target in the repetition sequence took 133 ms in one-distractor conditions and 200 ms in two-distractor conditions (Experiment 1), we expected RTs to increase accordingly. However, at least for one-distractor conditions, RTs were rather stable across the number of presentations. This might be due to two counteracting effects. On the one hand, the repetition sequences with few presentations of the target contained the least information to solve the task, and so detection was difficult and required the longest processing time. On the other hand, as more presentations of the target were added...
Fig. 4. Main results from Experiment 2. In (a), the mean percentage of repetitions detected in the rapid-serial-visual-presentation (RSVP) stream (left) and associated mean reaction time (RT; right) are plotted as a function of image-presentation rate. RT of 0 is the onset of the second presentation (i.e., the first repetition) of the target. In (b), mean identification performance (left), RT (middle), and confidence rating (right) in the memory task are shown. For readability, the presentation rates are displayed with logarithmic spacing. In the graph showing identification performance, the horizontal black line indicates the 25% chance level. Each data point represents the mean for 1 participant and condition, and the solid lines represent the mean across participants. The shaded areas are 95% confidence intervals.
Memory for Repeated Images

and the repetition sequence became easier to detect, the duration of the repetition sequence increased, also adding to RT. Following this line of reasoning for the two-distractor conditions, we hypothesize that the slight increase in RT with the number of presentations indicates that here the added stimulus duration of each presentation was longer than the processing time saved by the accompanying information gain. Already with only two presentations, RTs were faster with one distractor than with two—detection of the first repetition took longer when preceded by two compared with one distractor. The overall performance in the two-distractor conditions than in the one-distractor conditions might be due to the way the image was more efficiently masked before reappearing in the two-distractor conditions. An interesting observation can also be made about the link between the RTs and the onsets of the last target presentation in each condition (diagonal dashed lines in the right-hand graph of Fig. 2a): The curves intersect close to the 7-presentations condition, or around the point at which ceiling performance was reached. This is yet another indication that ceiling performance was reached. This is yet another indication that Presentations 8 through 10 of the targets were typically not beneficial for repetition detection or memory: If the repetition had not been spotted after 7 presentations, it probably would not be after 8, 9, or 10 presentations either. Further, we observe in the corresponding graph for Experiment 2 (right-hand graph in Fig. 4a) that at the lowest image-presentation rates, participants typically did not need all the repetitions of a target to complete the task. For example, at 4 Hz, all participants responded on average well before the fifth presentation of the target, meaning that often 4, or even 3, presentations were enough for detection.

An interesting comparison can be made between our results and those of Martini and Maljkovic (2009), who investigated memory for images presented once or twice in RSVP streams of up to 12 items presented at rates up to 77 Hz. In their design, all but the first and last images of the stream were targets in a subsequent new/old discrimination task. The targets that were presented twice were interspaced with one, two, or three other images. Discrimination was better for images that had been presented twice compared with once, and there was no effect of the spacing between the two presentations of a repeated image. Further, performance decreased with increasing presentation rate. This led the authors to conclude that the results are in agreement with the “total-time hypothesis,” which postulates that memory strength depends on the total exposure time. In the present study, we found an efficient encoding of repeated images that is not simply a result of the total time of exposure. For example, the number of distractors seems to be important both for repetition detection and memory encoding, even with identical total exposure time of the target. This has also been demonstrated in the auditory domain: Kaernbach (2004) found that repetition detection for noise segments 50- to 200-ms long was easiest for back-to-back presentation and that performance decreased with the amount of nonrepeated

![Image](image_url)

**Fig. 5.** Mean identification performance (left) and confidence in identifications (right), separately for detected and missed repetition targets in Experiment 2. Identification performance was measured as the percentage of correctly identified targets. For readability, the image-presentation rates are displayed with logarithmic spacing. The horizontal black line in the left-hand graph indicates the 25% chance level. In both graphs, each data point represents the mean for 1 participant, and the curved lines represent the mean across participants. The shaded areas are 95% confidence intervals.
distractor noise between repetitions (between 0 and 750 ms). Further, we found largely stable performance for image-presentation rates between 15 and 120 Hz, despite that fact that the total presentation time of the repeated image decreased over this interval from 333 ms to 42 ms (i.e., by a factor of 8). We argue that the repetition events are a key factor for memory encoding.

The apparent discrepancy between our conclusions and those of Martini and Maljkovic (2009) might be explained by several major differences in the experimental designs. In our experiments, repetition targets appeared at random time points and thus could not be anticipated; in Martini and Maljkovic’s (2009) study, all seven images occurring between the first and last images of the stream were targets on every trial. Further, in their study, the repetitions were not interleaved by distractors but by other targets, which could potentially have influenced the processing of the images. The same manipulation has been shown to diminish the attentional-blink effect where the detection of a first target impedes the detection of a second one appearing soon thereafter (Di Lollo, Kawahara, Shahab Ghorashi, & Enns, 2005; Nieuwenstein & Potter, 2006). Last, in our paradigm, the task was explicitly repetition detection.

If we rely on theories of periodic sampling of perception and attention, for which rhythmicity at different frequencies (mainly between 4 and 10 Hz) has been proposed (e.g., Dugué, McLelland, Lajous, & VanRullen, 2015; Fiebelkorn, Saalmann, & Kastner, 2013; VanRullen, 2016), we might expect nonmonotonic dependencies on the presentation rate of either the image stream or the target appearances in our paradigm. Here, we found no such clearly emerging “sweet spot” with better performance for specific presentation rates.

The lack of any visible effect of the gap between images in Experiment 2 is in line with the findings of previous work and with the idea that an image lingers perceptually until it is masked by the next image in the stream (Intraub, 1980; Potter, Staub, & O’Connor, 2004), an effect that has also been demonstrated in the anterior region of the superior temporal sulcus in monkeys (Keysers, Xiao, Foldiák, & Perrett, 2005). Hulme and Merikle (1976) showed that for briefly presented (50–300 ms) natural scenes that were masked immediately or after a short delay, recognition accuracy was not a function simply of the presentation time of the image but of the time from the onset of the image until the onset of the mask (total processing time). Similarly, in our data, we found no apparent influence even at high presentation rates, when the gap duration was long compared with the image-presentation time.

Identification performance in the memory task was highest for targets that had been detected as repeating, but in Experiment 1 (and to some extent in Experiment 2), performance was well above chance even for missed targets, suggesting that clear perception of the repetition might not be strictly necessary for memory formation. Confidence levels for identification of the missed trials reflected performance, indicating that participants were aware of recognizing these targets.

Is there a recency effect, in which the last repetition targets in the RSVP stream are the easiest to remember, or a primacy effect, in which increased memory strength is evident for the first targets in the stream? Although this study was not designed for testing trial-order effects, it is interesting to note that they were small overall (Figs. S3, S4, S9, and S10 in the Supplemental Material). In addition, there were only minor differences between short and long delays between the appearance of the same target in the RSVP stream and the memory task (Figs. S5 and S11 in the Supplemental Material). This observation was largely confirmed by Bayesian t tests (for details, see https://osf.io/dfhru/). Seemingly, any forgetting or lack of encoding happened to a similar extent for all targets regardless of the number of trials intervening between repetition and prompting of a target, on a time scale from several seconds to minutes. Because this analysis has shortcomings (i.e., unbalanced number of trials and contributions from different conditions), further studies will be needed to confirm this effect and to map the time course of the (presumed) fading from memory of the repeated images. Previous research has revealed trial-order effects on target retention after RSVP streams, namely forgetting as a function of the number of intervening test items and of elapsed time on the time scale of seconds. However, such research has found no effect of presentation order, except possibly for the first and last picture (Potter & Levy, 1969; Potter et al., 2004; Potter, Staub, Rado, & O’Connor, 2002). Given that in the auditory domain, even meaningless noise snippets have been found to induce memory traces lasting for days or even weeks after repetition (Agus et al., 2010), it is not unreasonable to expect some lasting memory for the repeated images in our paradigm.

A related question is whether some images are more likely than others to be detected and remembered (Isola, Xiao, Parikh, Torralba, & Oliva, 2014). Broers, Potter, and Nieuwenstein (2018) showed that images that had previously been identified as memorable were better remembered than nonmemorable images when presented once in short RSVP sequences of between 2.8 Hz and 77 Hz (13- to 360-ms presentation time per image) and that performance for the memorable pictures increased with presentation time across the entire explored range. For the nonmemorable pictures, this increase was seen only for presentation rates below 8.3 Hz (presentation times over 120 ms). Because of the
nature of our paradigm, in which images used as targets varied between participants and conditions, we could not easily explore the role of image content.

On the basis of subjective reports, we hypothesize that two rather distinct strategies are deployed, depending on the image-presentation rate: At low rates, the content of each picture is readily available and can be processed on a cognitive level (e.g., naming each image). As the presentation rate increases, cognitive strategies become less practical, and a more automatic way of detecting repetitions, associated with a perceptual “shining through” of the target, takes over instead.

What memory mechanisms allowed our participants to identify the repeated images? There are several arguments against a pure working memory account. First, there were only a few seconds between the repetition targets, and the detection task was highly attention demanding, leaving little room for active storage and mental repetition of the targets. In addition, learning an ordered sequence of labels for the repeated images was limited because of the randomized order of the targets in the memory task and because many of the images did not have easily labeled content. Second, even missed repetition targets were correctly identified in the memory task, which suggests that encoding was partly nonconscious. Third, in the exploratory analysis of trial-order effects, we found no strong influence of the duration (and the number of other intervening items) between repetition and identification of an image on identification performance.

We speculate that the encoding of lasting memories for repeating stimuli is a natural and fundamental function of the brain. Memory formation for anything that repeats in the environment allows for fast processing of commonly occurring items, a feature that has a clear ecological value in interpreting and predicting sensory input in any modality. Neural mechanisms based on STPD-like learning that make neurons selective could sustain this type of memory encoding (Bichler et al., 2012; Masquelier, 2018; Masquelier et al., 2008). Was neural selectivity created during our experiments? This possibility could theoretically be tested by recording neurons during stimulation. In the auditory domain, recent work has revealed rapid formation of sharp selectivity to repeated noise snippets, as measured by event-related-potential responses (Andrillon et al., 2015). These memory-evoked potentials were generated automatically even though the task was not to detect reoccurring targets and even when participants’ attention was diverted. These findings, together with the impressive capacity for repetition detection and memory demonstrated in the present experiments, suggest the presence of an efficient, low-level neural mechanism for processing repetitions.

Action Editor
D. Stephen Lindsay served as action editor for this article.

Author Contributions
E. Thunell and S. J. Thorpe developed the study concept and design. Data were collected and analyzed by E. Thunell, and both authors interpreted the results. The manuscript was drafted by E. Thunell, and S. J. Thorpe provided critical revisions. Both authors approved the final version of the manuscript for submission.

Acknowledgments
We thank Céline Cappe for lending us her equipment and Greg Francis and Jean-Michel Hupé for useful discussions.

Declaration of Conflicting Interests
The author(s) declared that there were no conflicts of interest with respect to the authorship or the publication of this article.

Funding
This work received funding from the European Research Council (ERC) under the European Union’s Seventh Framework Programme (FP/2007-2013)/ERC Grant Agreement No. 323711 (M4 project).

Open Practices
Neither of the experiments reported in this article was formally preregistered. The modified images from the ImageNet database training set are available on the Open Science Framework at https://osf.io/7epb/. The data and the details of the Bayesian analysis are available at https://osf.io/fdhnu/.

References


