

Cognition and Behavior

# Perceptual Learning with Complex Objects: A Comparison between Full-Practice Training and Memory Reactivation

Chiu-Yueh Chen<sup>1,2</sup> and Hans Op de Beeck<sup>1,2</sup><https://doi.org/10.1523/ENEURO.0008-19.2021><sup>1</sup>Brain and Cognition, Katholieke Universiteit Leuven, Leuven 3000, Belgium and <sup>2</sup>Leuven Brain Institute, Katholieke Universiteit Leuven, Leuven 3000, Belgium

## Abstract

Perception improves with repeated exposure. Evidence has shown object recognition can be improved by training for multiple days in adults. Recently, a study of Amar-Halpert et al. (2017) has compared the learning effect of repetitive and brief, at-threshold training on a discrimination task and reported similar improvement in both groups. The finding is interpreted as evidence that memory reactivation benefits discrimination learning. This raises the question how this process might influence different perceptual tasks, including tasks with more complex visual stimuli. Here, this preregistered study investigates whether reactivation induces improvements in a visual object learning task that includes more complex visual stimuli. Participants were trained to recognize a set of objects during 5 d of training. After the initial training, a group was trained with repeated practice, the other a few near-threshold reactivation trials. In both groups, we found improved object recognition at brief exposure durations. Traditional intense training shows a daily improvement; however, the group with reactivation does not reach the same level of improvement. Our findings show that reactivation has a smaller effect relative to large amounts of practice.

*Key words:* consolidation; object recognition; perceptual learning

## Significance Statement

Perceptual learning helps to explore adult plasticity in visual processing. Gradual improvements in the perception of complex objects have been demonstrated across multiple daily training sessions of hundreds of trials. These improvements in the trained objects and the transfer to new objects, in that sense, support “practice makes perfect.” Recent research challenges this idea, and suggests that a few critical reactivation trials can boost the learning processes. Here, we extend this idea to other learning tasks and investigate the extent to which short reactivation with a small number of trials can replace extensive training with complex visual objects. In our paradigm, we found larger training effects with extensive training.

## Introduction

Sensory information processing can be improved, but improvement requires repetitive practice. Repeated presentations of the same stimuli induce perceptual learning. A substantial body of evidence demonstrates training-induced improvements in the perception of relatively simple visual aspects, such as frequency discrimination (Fiorentini and

Berardi, 1981), orientation discrimination (Aberg et al., 2009; Jeter et al., 2010), and feature discrimination (Karni and Sagi, 1993; Censor et al., 2006; Censor and Sagi, 2008; Amar-Halpert et al., 2017). Interestingly, similar learning curves have been reported with more complex objects (Furmanski and Engel, 2000; Baeck et al., 2012, 2014, 2016; Van Meel et al., 2016). This so-called object learning

Received August 18, 2020; accepted January 30, 2021; First published February 5, 2021.

The authors declare no competing financial interests.

Author contributions: C.-Y.C. and H.O.d.B. designed research; C.-Y.C. performed research; C.-Y.C. and H.O.d.B. analyzed data; C.-Y.C. and H.O.d.B. wrote the paper.

involves improved recognition of objects after multiple days of training. This improvement in object recognition under perceptually challenging conditions seems related to activity in high-level object-selective cortex (Grill-Spector et al., 2000; Van Meel et al., 2016).

While the boundary between simple and complex processing is difficult to draw, together these findings demonstrate that a wide variety of visual capabilities improve with extensive trainings. Recently however, in a texture discrimination task, the same profile of learning across days was observed using only a limited number of reactivation trials in the subsequent training days (Amar-Halpert et al., 2017). This study made use of a hypothetical window of opportunity during memory consolidation, where memories are re-evoked in the subsequent days after the initial encoding training. This reactivation needs only a few trials but results in similar behavioral improvements compared with the usual practice with hundreds of training trials. A similar mechanism in the domain of motor learning suggests memory consolidation because of reactivation in the context of a finger-tapping task (Walker et al., 2002, 2003). In that case, the induced performance benefits occurred in a brief reactivation which is <60 s (de Beukelaar et al., 2014). Although these training paradigms have shown stable and long-lasting effects in simple motor and visual learning with such short reactivations, no study has explored if the same phenomenon can be observed in more complex processing. The number of studies on this phenomenon is very limited and it is important to explore the necessary and sufficient conditions of these effects.

Another aspect to consider in visual learning is learning selectivity. Indeed, the training-induced learning effects are observed specifically for the stimuli set used for training. However, the degree of generalization (or transfer) of learning to new (but related) stimuli varies. In fact, the learning effects can show a lack of generalization to other objects (Furmanski and Engel, 2000). On the other hand, studies have typically noticed a partial generalization across stimulus size (Furmanski and Engel, 2000) and unseen images of the same objects (Baeck et al., 2016). It is unclear how the aforementioned reactivation protocol would interact with specificity and generalization.

This preregistered study will therefore test (1) whether the short reactivation strategy can induce object learning to the same extent as classical intensive training and (2) how selective this learning process is. The experimental design will in many details be based on earlier object learning experiments (Baeck et al., 2012), using the same materials and dependent variable they used. In addition, we will add a between-subject manipulation of reactivation

training similar to the procedure from Amar-Halpert et al. (2017). Based on the previous success with this reactivation protocol, we predict similar learning curves in this protocol compared with the standard object training protocol with a much higher number of trials.

Earlier studies on object recognition training have started with a group of common objects as stimuli. With hundreds of repetitive trials, recognition can be improved by training. However, we hope to uncover the detailed reasons for changes in object recognition. Now, our study aims to bridge the gap between object learning paradigms that involve long exposure and shorter paradigms that rely on memory reactivation. This work will inform us about the mechanisms underlying learning-induced improvements in object recognition. The larger the extent that memory reactivation is involved in such improvements, the more improvement we expect to find after brief periods of training that are designed to reactivate memory.

## Materials and Methods

The approved Stage 1 protocol, the anonymized study data and digital materials can be found on the Open Science Framework (OSF) at <https://osf.io/utx6n/>.

### Participants

Fifty-two participants in this study were randomly assigned to two groups: 26 in the full-practice groups (aged  $22.5 \pm 5.2$  years, six males and 20 females) and 26 in the short-reactivation group (aged  $22.1 \pm 4.3$  years, two males and 24 females).

Participants of either gender (aged between 18 and 40) were recruited online through a university online recruitment system (SONA), Facebook, as well as through banners and leaflets. The volunteers received monetary rewards.

Initially, a planned sample of 50 was set from a power analysis over 0.90. Data collection was prone to no-shows, which is why we scheduled >50 participants. In the last week of testing, we needed one participant but scheduled three and all three showed up. As such a final sample of 52 was obtained.

There was no a priori limit on the proportion of male/female participants. The experiment was approved by the Social and Societal Ethics Committee of Katholieke Universiteit Leuven (G-2017 121045). Participants signed an informed consent before every session. The following criteria was used to exclude a participant's data: (1) a participant does not attend all sessions; (2) the obtained threshold value on the first or the last day is worse than the baseline (>120 ms).

On top of the final sample of 52, three more participants started with the first session, but were removed from the data file; one because of the obtained threshold value on the first day (140.94 ms) was worse than the starting value of 120 ms set a priori, one because of cancelation of the later slots, and one because of being outside of the pre-defined age range.

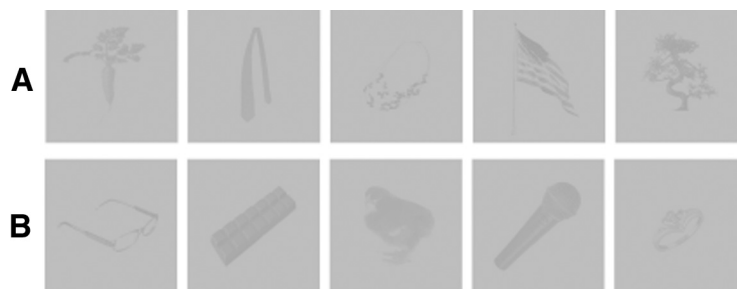
This work was supported by Katholieke Universiteit Leuven Research Council Grant C14/16/031.

Correspondence should be addressed to Hans Op de Beeck at [hans.opdebeeck@kuleuven.be](mailto:hans.opdebeeck@kuleuven.be).

<https://doi.org/10.1523/ENEURO.0008-19.2021>

Copyright © 2021 Chen and Op de Beeck

This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International license, which permits unrestricted use, distribution and reproduction in any medium provided that the original work is properly attributed.



**Figure 1.** Two stimulus sets. Each set contains five different object categories. Panels **A** and **B** display 2 of the 6 subsets used in the experiment.

## Apparatus

The whole experiment was conducted using a Dell desktop computer (GX-780), using MATLAB and Psychtoolbox 3 (Brainard, 1997; Pelli, 1997; Kleiner et al., 2007). Visual stimuli were displayed on a 16-inch CRT monitor (Dell 790) with a  $1024 \times 768$  pixels resolution at 100 Hz. The room was dim and viewing distance was 90 cm.

## Stimuli

The full stimulus set consisted of 40 gray-scale pictures of common manmade and natural objects that were used in a previous study (Baeck et al., 2012). The contrast of stimuli was reduced to 12.5% of the original contrast to make objects harder to recognize. Masking stimuli are made by a combination of fragments ( $70 \times 70$  pixels) of all different object pictures. Image size was  $450 \times 450$  pixels (8.7 visual degrees). All stimuli were  $\gamma$  corrected to create a linear luminescence range. As the  $\gamma$  correction decreased overall contrast, an inverse  $\gamma$ -correction was applied to the masking stimuli to increase the contrast of the masks and thus obtaining a more robust masking effect.

### Selection of image set

In Baeck et al.'s (2012) work, the large stimulus set was divided in two subsets. One subset was used for training the other subset as a control. This design assumed that 20 stimuli per subset would be sufficient to average out possible stimulus-specific variations in difficulty (e.g., a stimulus that is more difficult than other stimuli). However, the large number of stimuli results in less exposure to each individual stimulus. In the current study there is an additional disadvantage, namely that the number of reactivation trials goes up linearly with the number of stimuli. For that reason, we decided to limit the number of stimuli in each subset to five stimuli.

We partitioned the stimuli in six subsets of five stimuli, then tested five pilot participants with six two-down, one-up staircases per subset. The obtained thresholds of these pilot trials helped us to select a number of stimulus subsets with equal difficulty. However, the task with five stimuli became so easy that the responses were achieved at the minimum stimulus duration without fluctuation, making it impossible to investigate the main effect. To prevent visual adaptation, the stimulus size was changed randomly from 250 pixels to 450 pixels, and this additional

need for size invariance increased the difficulty of the recognition at short stimulus durations.

Two stimulus sets (Fig. 1) were counterbalanced across participants so that all stimulus sets will be included to the same extent as training and control stimuli between the two subject groups.

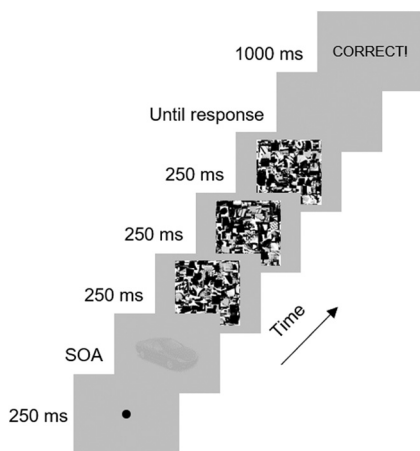
## Object learning task

Each trial started with a fixation and the stimulus presentation for a variable time [stimulus onset asynchrony (SOA)]. Next, three consecutive masks were presented at the same location for 250 ms each, to prevent further visual processing (Op de Beeck et al., 2007). The presented position of stimuli and masks was randomized with a maximum deviation of 1.8 degrees from the screen center, the size-invariant stimuli and masks ranging randomly from  $250 \times 250$  pixels to  $450 \times 450$  pixels. The variable stimulus duration was determined through two interleaved two-down, one-up staircases. Stimulus duration was initially set at 120 ms (12 frames at a 100-Hz refresh rate), decreased by 10 ms (one frame) after two consecutive correct answers and increased by 10 ms after each incorrect answer. Participants were requested to type the first two letters of the name of the presented object. Three-letter responses have been used before (Baeck et al., 2012), but we can simplify this to two letters because stimulus sets included only five stimuli. A “true” or “false” feedback was shown after each trial. In case of a wrong answer, the correct object name was provided (see Fig. 2).

Note that the procedure of shortening stimulus duration as the SOA gets smaller in could arguably result in a situation in which challenges in recognition performance are because of reductions in perceived stimulus contrast in low-level visual areas. We cannot rule out this possibility, yet it is assuring that the same manipulation was used in previous studies that found correlations with object recognition performance in object-selective cortex (Grill-Spector et al., 2000) and causal effects after stimulation in lateral occipital cortex (Van Meel et al., 2016).

## Procedure

Participants were randomly assigned to two training paradigms (full-practice vs short-reactivation trainings). Each participant was trained with one subset of five



**Figure 2.** Example of one trial. The durations between the object and the mask stimuli (SOA) are determined based on an adaptive interleaved two-down, one-up staircase procedure. The object stimulus in this example trial is a car (“auto”). If the participant types “au,” the feedback “CORRECT!” (English: “CORRECT”) is shown. If the participant gives another response, then “FOUT! HET was auto” (English: “WRONG! It was car”) appears.

stimuli. Threshold values are the average of the last four reversals of all staircases per day.

One other subset of five stimuli was used as a control set for individual participants, only to be seen during the first and last session. Across participants, each stimulus subset served equally often as a training set and as a control set.

The participants in the full-practice group completed standard training sessions between the first and the last day (see Table 1). The other participants in the short-reactivation group completed a standard training session on the first day and perform only short reactivation on the other 3 d. The fifth day is a test session which is the same for both groups.

*Preview*

In earlier work (Furmanski and Engel, 2000; Baeck and Op de Beeck, 2010; Baeck et al., 2012, 2014, 2016; Van Meel et al., 2016), each session started with a preview of all stimuli presented in that session. A preview of the

stimuli (2 s each) with their corresponding names was displayed to ensure that participants know the object images and their names. In the present experiment, we included this preview in the first and the last session for all 10 object stimuli (two subsets). The preview was not shown in the intermediate training sessions to not interfere with the reactivation procedure.

*First test and training session*

The first session involved a preview and eight experimental blocks of 100 trials (800 trials). Each block comprised two interleaved staircases of 50 trials. Baeck et al. (2012) only used 40 trials per staircase, but we expect to be able to collect more trials per unit of time because the required responses have been simplified (two instead of three letters). Each block included only five stimuli, either the stimuli that were trained or the control stimuli. Of the eight blocks in the first session, four blocks included trained and four blocks control stimuli, interleaved, and the stimulus set that came first was counterbalanced across participants. The total duration of the experimental session on day 1 would last 1 h.

*Training session*

A standard training session involved eight experimental blocks of 100 trials (800 trials) with each block comprising two interleaved staircases of 50 trials. Standard training sessions only included the trained stimuli, the control stimuli were not shown.

*Reactivation session*

Participants performed five near-threshold trials of each stimulus, resulting in 25 reactivation trials per session. Threshold values were the average of the last four reversals of the last four staircases on day 1 for that participant.

*Final test session*

The test session on the final day was the same as the first half of the first session, with a preview of the 10 object images and four blocks of 100 trials. This test session was shorter than the first session, because it only served to assess the thresholds and not to induce a large amount of training.

**Table 1: Study design**

Full-practice training paradigm				
Time	Day 1	Days 2–4	Day 5	
Session	First test and training	Training	Final test	
Training trials	400	800	200	
Control trials	400		200	
Short-reactivation training paradigm				
Time	Day 1	Days 2–4	Day 5	
Session	First test and training	Reactivation	Final test	
Training trials	400	25	200	
Control trials	400		200	

Procedure of the adapting training sequence for the full-practice training paradigm and the short-reactivation training paradigm.



**Table 2: Summary of statistical analysis**

	Data structure	Type of test	Power/CIs
a	Normal distribution	Independent <i>t</i> test	0.98
b	Normal distribution	Independent <i>t</i> test	0.07
c	Normal distribution	Independent <i>t</i> test	0.09
d	Normal distribution	Repeated measures ANOVA	1
e	Normal distribution	Dependent <i>t</i> test	95% CI [12.35,17.80]
f	Normal distribution	Dependent <i>t</i> test	95% CI [8.78,14.63]
g	Normal distribution	Independent <i>t</i> test	0.73
h	Normal distribution	Independent <i>t</i> test	0.76
i	Normal distribution	Bootstrap <i>t</i> test	95% CI [-15.54,-2.70]
j	Normal distribution	Independent <i>t</i> test	0.67
k	Normal distribution	Dependent <i>t</i> test	95% CI [7.66,14.17]
l	Normal distribution	Dependent <i>t</i> test	95% CI [7.17,11.96]
m	Normal distribution	Independent <i>t</i> test	0.17
n	Normal distribution	Independent <i>t</i> test	95% CI [2.34,6.79]
o	Normal distribution	Independent <i>t</i> test	95% CI [-1.81,4.55]

### Analysis pipeline

Learning-related changes were quantified by the threshold values across sessions, which is the most obvious manner to characterize performance when using an adaptive procedure (psychometric curve fitting is difficult given the uneven sampling of this function, see [Baeck and Op de Beeck, 2010](#)). For each participant, threshold values were the average of the last four reversals of every staircase complete with a stimulus set in a session. A lower threshold value corresponds to better performance. If the obtained threshold on day 1 or day 5 is worse than the baseline (120 ms), the individual data will be excluded. The data were presented as mean and the SEM of two stimulus sets (trained, control), two training paradigms (full-practice, short-reactivation) and time (sessions). A decline in thresholds over sessions indicates a training effect.

Before analyzing the training effect, individual data on the initial session was checked for equivalence between the two training paradigms. Next, the main analysis evaluated learning effect with *t* tests in line with [Baeck and Op de Beeck \(2010\)](#); [Baeck et al. \(2012, 2014, 2016\)](#) and [Van Meel et al. \(2016\)](#). In order to test the central hypothesis of the current paper, the final session performance in the two groups was compared with find out whether the full-practice group has reached lower thresholds compared with the short-reactivation group. We further tested the specificity for the control stimuli among different training groups. Results are reported with *p* values, confidence interval (CI), and effect size (see [Table 2](#)).

### Equivalence test for comparing the groups before training

To confirm that the observed training effects are meaningful when assessed from the performance in the last day, we evaluated whether the two groups had equal performance on the first session using an independent *t* test. We did not run an equivalence test as proposed originally because we did not set the bounds of the latter a priori. Irrespective of the outcome of this analysis, we also report the results when data were normalized for the performance on day 1. The same overall conclusions are reached with and without normalization, although quantitatively the numbers change.

### Overall training effect

The main analysis assessed the effect of training between day 1 and day 5, with two paired *t* tests for each training paradigm.

### Group comparison in terms of training effect

We compared the two groups in two ways. First, we compared the day 5 performance between the full-practice and the reactivation group with an unpaired *t* test. Second, as done by [Amar-Halpert et al. \(2017\)](#), we would estimate the level of improvement. The training effect is sometimes referred to as a learning rate. Here, to use a consistent terminology relative to the study of [Amar-Halpert et al. \(2017\)](#), we also use the term learning rate. The learning rate would be computed that divided the differences of day 1 and a target day by day 1 performance, for example (day 1 – day 5)/day 1, multiplied by 100 to obtain percentages.

### Specificity of object learning

Following the analysis of the overall learning effect, we investigated the specificity of the training in days 2–4 to the untrained stimuli. Two paired *t* tests were conducted to compare the control stimuli performance between the first and final test sessions within a group of subjects. An unpaired *t* test was conducted to compare the day 5 performance between the full-practice and the short-reactivation group.

### Effect and sample size calculation

The key comparisons made by [Amar-Halpert et al. \(2017\)](#) in the 5-d standard practice, memory-reactivation, and 2-d standard practice suggest that the brief reactivations during training do improve discrimination thresholds. The original paper did not provide the raw and average values at the test and retest sessions; we estimated the pooled SD to be 20% within two groups from the given SEM of 5.9% and 5.5% in each group. Here, we present effect sizes and required sample sizes, calculated with a data analytics software (G\*Power 3.1.9.2; RRID: [SCR\\_013726](#)).

Amar-Halpert et al. (2017) reported that learning rates in the two groups ranged from 20.6% (memory reactivation) to 26.6% (standard practice), with a nonsignificant difference ( $F_{(1,22)} = 0.56$ ,  $p = 0.46$ ) between total learning rates in the standard-practice and the memory-reactivation groups.

Despite a relatively low number of participants per group ( $N = 12$ ), the study of Amar-Halpert et al. (2017) had a reasonable power to detect a learning effect in a group, because these effects are large. To find an effect of 20.6% with SD of 20%, the power is 0.75. If our null hypothesis is that memory-reactivation would result in no learning effect at all and the alternative hypothesis states that there is as much learning as in the standard practice group, then with  $N = 18$  we would have a power of 0.90. However, we could also hypothesize that the reactivation might result in some learning, albeit much smaller than in the standard-practice group. Thus, to safeguard us against this possibility, we opt for a sample size of  $N = 25^a$  in each group, which is double the number in the original study.

## Results

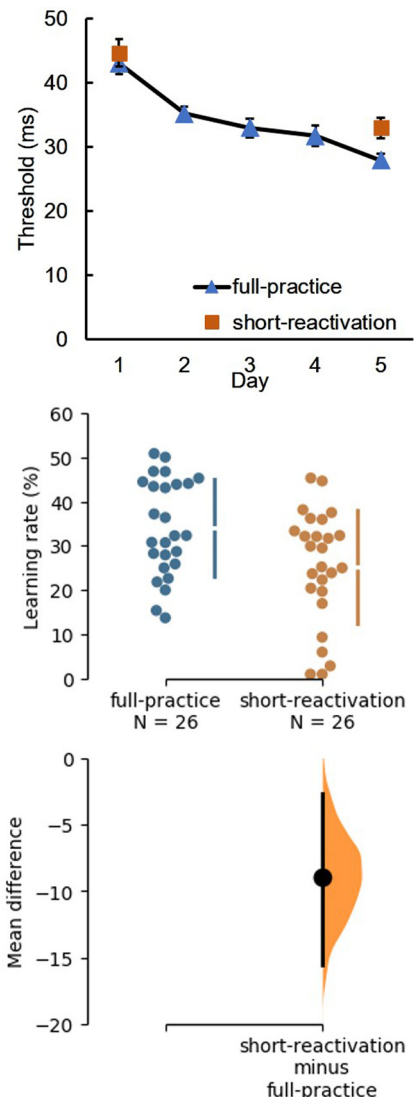
All data and analysis scripts are publicly available on the OSF (<https://osf.io/utx6n/>). This study obtained 52 participants, 26 in the full-practice group and 26 in the short-reactivation group. Participants in both groups performed a 5-d training with the same amount of trials on the first and final days. Participants in the full-practice group performed a standard training session of 800 trials in three daily training sessions (days 2–4). Participants in the short-reactivation group performed 25 at-threshold trials in three daily reactivation sessions.

Threshold values, representing the performance per subject, are the average of the last four reversals of every staircase. The reversals include local maximums and local minimums. The local maximums are calculated from a function (findpeaks) in MATLAB. In order to perform this analysis of the local minimums, the data are multiplied by  $-1$ .

The initial performance on full-practice and short-reactivation training is tested for equivalence. Two-tailed  $t$  tests confirmed that both groups had similar performance on the first day. Using the data of both stimulus sets, there was no significant difference,  $t_{(50)} = 0.4569$ ,  $p = 0.5793$ ,  $d = 0.1267^b$ , nor was there a difference when tested only on the thresholds for only the trained stimulus set,  $t_{(50)} = 0.6063$ ,  $p = 0.5471$ ,  $d = 0.1682^c$ .

Although the difference is not significant, there is a trend toward higher initial thresholds in the reactivation group. Importantly, all our conclusions are not only backed up by day 5 performance, but also by analyses of the learning rates that take into account the baseline performance at day 1.

During the reactivation sessions, subjects were given brief training with the near-threshold trials. To estimate the individual threshold for each subject, the final four reversals of last four staircases on day one were averaged, mean  $\pm$  SE:  $40.75 \pm 1.86$  ms. Based on this, the stimulus duration used for the reactivation trials was  $41.54 \pm 2.05$  ms (note that per participant the duration is a multiple of 10). The average accuracy during the

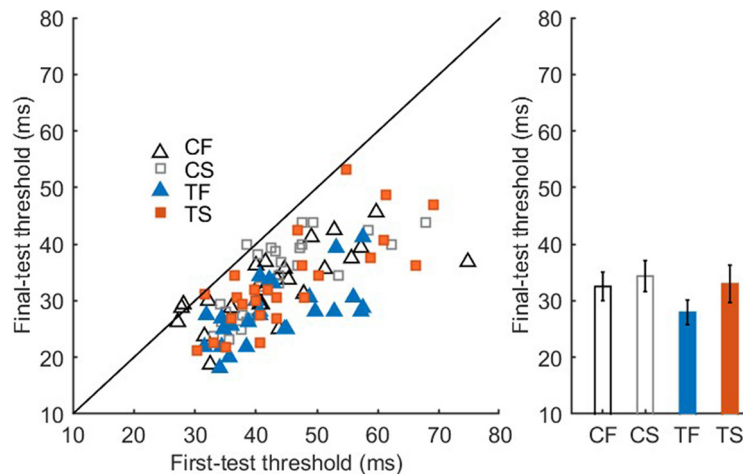


**Figure 3.** Overall learning effect in the full-practice group and the short-reactivation group for the trained stimuli. Top panel, Performance thresholds are plotted as a function of time and group. Error bars represent the SEM. Middle panel, The distribution shows the spread of the learning rates in each group. Circles represent the learning rates of individual participants. The gapped lines represent the SD of each group. Bottom panel, A bootstrapped resample distribution depicts the learning rate difference between two groups. A circle represents the difference between two groups of  $-8.90$  and the end of the vertical black bars represent 95% confident interval of  $-15.53$  and  $-2.70$ ,  $p = 0.0072$ .

reactivation trials was  $60 \pm 3.86\%$  on day 2,  $67.54 \pm 4.03\%$  on day 3, and  $70.62 \pm 4.29\%$  on day 4.

## Learning effect

After testing the initial performance between two groups, the performance on trained stimuli after training was assessed with two paired  $t$  tests. Mean thresholds for both groups are shown in Figure 3. A learning effect was found for the full-practice group as in previous studies



**Figure 4.** Recognition thresholds in the full-practice group and the short-reactivation group. Scatter plot of the thresholds of individual participants in the first and last day, plotted separately per stimulus set and per group ( $N=26$  in each group). The diagonal line represents the same performance on the first and last day, all points falling under this line represent a better performance on the last day compared with the first day. CF, control stimuli in full-practice group; CS, control stimuli in short-reactivation group; TF, trained stimuli in full-practice group; TS, trained stimuli in short-reactivation group. Error bars represent 95% CI.

( $F_{(4,120)} = 56.87$ ,  $p < 0.001$ ,  $\eta^2 = 0.70^d$ ). Overall, performance thresholds were lower on day 5 than the day 1 in the full-practice group ( $t_{(25)} = 11.4$ ,  $p < 0.001$ ,  $d = 2.24^e$ ), as well as in the short-reactivation group ( $t_{(25)} = 8.24$ ,  $p < 0.001$ ,  $d = 1.62^f$ ).

Then we compared the learning-induced changes for the trained stimulus set between two groups, as inferred from the performance threshold on the final day and the learning rate. The a priori hypothesis was that participants in the short-reactivation group and in the full-practice group improved equally. However, participants in the short-reactivation group were found to perform less well on the final test session (day 5) than participants in the full-practice group,  $t_{(50)} = 2.6096$ ,  $p = 0.0119$ ,  $d = 0.72^g$ . A similar finding was found when focusing on the learning rate. The learning rates for full-practice (Mean  $\pm$  SD,  $34.13 \pm 10.93\%$ ) and for reactivation group ( $25.23 \pm 12.74\%$ ) were significantly different ( $t_{(50)} = 2.7047$ ,  $p = 0.0093$ ,  $d = 0.75^h$ ). For the key finding, we compared the mean difference of the learning rate between groups and performed a bootstrapped distribution by 5000 resamples. The difference of learning rates in two groups was  $8.90\%$  with a 95% CI from  $-15.54$  to  $-2.70^i$ .

Note that these comparisons still overestimate the effect of the reactivation sessions. Much of the improvement compared with day 1 might be a consequence of the many trials in day 1. Motivated by the surprising result of much smaller learning in the reactivation group, we performed an additional analysis that was not mentioned in the preregistration. We questioned whether there would be any beneficial effect of the reactivation trials if we take into account the training induced by day 1. We compared the learning rate at day 5 in the reactivation group, with the learning rate at day 2 in the full-practice group. These two cases are comparable in the sense that participants have received 1 day of training (day 1), with then the reactivation trials as an additional exposure for the reactivation

group. This test is not free of confounds though, in particular, the time intervals are not the same, which should mostly bias us toward a null effect (no added benefit of the reactivation trials because potentially undone by a longer time interval). When tested, day 2 learning rate in the full-practice group ( $17.07 \pm 11.14\%$ ) was lower than the day 5 learning rate in the reactivation-related changes,  $t_{(50)} = 2.46$ ,  $p = 0.0175$ ,  $d = 0.68^j$ . This finding suggests that the reactivation had some beneficial effect, but, given our earlier tests, much less than a full practice protocol.

### Specificity

Following the analysis of the learning effect, we investigated the specificity of learning to the trained stimulus set by analyzing the thresholds for the control stimuli that were only present at the first and final test sessions. The threshold for the control stimuli was lower on the final test session (day 5) than the first test session (day 1) in the full-practice group,  $t_{(25)} = 6.9011$ ,  $p < 0.001$ ,  $d = 1.35^k$ , and in the short-reactivation group,  $t_{(25)} = 8.2303$ ,  $p < 0.001$ ,  $d = 1.61^l$  (Fig. 4). On the final test session, performance for the control stimuli did not differ between groups,  $t_{(50)} = 1.0095$ ,  $p = 0.3176$ ,  $d = 0.80^m$ . Furthermore, on the final test session, participants in the full-practice group were better at recognizing trained stimuli than the control stimuli,  $t_{(25)} = 4.2261$ ,  $p = 0.0$ ,  $d = 0.83^n$ , while this was not the case in the short-reactivation group,  $t_{(25)} = 0.8877$ ,  $p = 0.3831$ ,  $d = 0.17^o$ .

### Discussion

The main objective of this study was to test the generality of previous work suggesting that in visual learning a short reactivation protocol results in as much learning as a more traditional time-intensive training. Our main finding is a significantly better learning effect in object recognition for participants who performed traditional repeated

training. Although there is still a small beneficial effect of the reactivation protocol, this improvement is much less than what is achieved through a time-intensive training. The lack of strong training effects of the reactivation protocol with one set of stimuli is probably also the main reason why there is no specificity of training when comparing performance between the trained stimuli and another set of stimuli that were not shown during reactivation.

This study successfully establishes perceptual learning and extends the prior knowledge of the effects of memory reactivation. Our findings of a limited effect of memory reactivation resonate with other domains in which it has proven hard to identify the boundary conditions that influence the beneficial effects of memory reactivation. In several domains there has been a discussion that resulted from a variety of experimental outcomes. The original formulation of memory reconsolidation theory comes from the domain of fear conditioning. Researchers have hypothesized the stabilized memory can be modulated and even destabilized by performing extinction training during the reconsolidation window (Monfils et al., 2009; Schiller et al., 2010). However, the effect of fear memory destabilization reported in the original studies could not always be replicated, sometimes the reactivation did not differ in or outside the hypothesized critical period, and in some studies there was even no effect of reactivation (Luyten and Beckers, 2017; Chalkia et al., 2020). Similar discussions have arisen in the domain of motor skill learning. Evaluated by obtaining a key typing task, new learning interfered with performance (Walker et al., 2003). The expected reconsolidation effect was absent in a direct and conceptual replication (Hardwicke et al., 2016; Walker and Stickgold, 2016). In the domain of perceptual learning, the positive findings of Amar-Halpert et al. (2017) are accompanied by reactivation effects in similar paradigms such as orientation discrimination (Bang et al., 2018). Overall, while a large literature also supports the effects of reactivation and its induced reconsolidation (for review, see Lee et al., 2017), a lot remains to be done to understand the boundary conditions under which a reactivation protocol is effective.

It is important to note that our study was not meant to directly replicate the study of Amar-Halpert et al. (2017). Instead, we wanted to investigate to what extent similar effects of reactivation could be found in a paradigm that focuses on the learning of more complex visual objects. We decided to stay close to the object naming paradigm that has been used in several previous studies (Furmanski and Engel, 2000; Baeck and Op de Beeck, 2010; Baeck et al., 2014, 2016; Van Meel et al., 2016), rather than trying to come up with an object recognition task that would be as similar as possible to the texture discrimination task of Amar-Halpert et al. (2017). As a consequence, there are many differences between the two protocols on top of the difference in domain (texture vs objects), and that might affect the results given that perceptual learning is sensitive to a lot of variables, including stimulus parameters (Sagi, 2011). To mention a few differences, we used a different method to obtain thresholds (an adaptive procedure, in contrast to the method of constant stimuli), we

found different threshold values (which could be because of the adaptive procedure, or to a different strength of masking), our paradigm included a preview of the stimuli which might facilitate performance before training, the stimulus-response mapping is more complex, the reactivation includes more trials because of having more stimuli, and we have a larger sample size which results in a higher power and thus a higher probability of finding group differences.

The improvements that we find because of learning are small in absolute magnitude, and so are the differences between groups. However, the differences are large in percentages, and with this paradigm we and others have consistently found highly replicable learning effects despite the small size in absolute terms (Furmanski and Engel, 2000; Baeck and Op de Beeck, 2010; Baeck et al., 2012). This object learning paradigm also results in very consistent and replicable effects. Nevertheless, we cannot exclude the possibility that our finding of less learning in the reactivation group would be related to specific properties of our methods, such as the way we estimate the thresholds or the limited temporal resolution with which we can adjust task difficulty (limited by the frame rate of the monitor).

The results of the current study support the claim that the previously noted improvements in the object naming task require large amounts of practice. In comparison to what was done in many previous studies (Furmanski and Engel, 2000; Baeck and Op de Beeck (2010); Baeck et al., 2012, 2014, 2016), the reactivation protocol is very short. It did not result in much additional improvement on top of the effect of the first day of extensive training. However, many experimental parameters might affect the strength of effects induced by reactivation, and we might simply not have found the optimal conditions. Thus, instead of concluding that reactivation has a smaller effect than the brute-force method of large amounts of practice, we hypothesize that its effects might simply depend more on the circumstances.

## References

- Aberg KC, Tartaglia EM, Herzog MH (2009) Perceptual learning with Chevrons requires a minimal number of trials, transfers to untrained directions, but does not require sleep. *Vision Res* 49:2087–2094.
- Amar-Halpert R, Laor-Maayany R, Nemni S, Rosenblatt JD, Censor N (2017) Memory reactivation improves visual perception. *Nat Neurosci* 20:1325–1328.
- Baeck A, Op de Beeck HP (2010) Transfer of object learning across distinct visual learning paradigms. *J Vis* 10:17.1–9.
- Baeck A, Windey I, Op de Beeck HP (2012) The transfer of object learning across exemplars and their orientation is related to perceptual similarity. *Vision Res* 68:40–47.
- Baeck A, Rentmeesters N, Holtackers S, Op de Beeck HP (2014) The effect of sleep in perceptual learning with complex objects. *Vision Res* 99:180–185.
- Baeck A, Maes K, Van Meel C, Op de Beeck HP (2016) The transfer of object learning after training with multiple exemplars. *Front Psychol* 7:1386.
- Bang JW, Shibata K, Frank SM, Walsh EG, Greenlee MW, Watanabe T, Sasaki Y (2018) Consolidation and reconsolidation share



- behavioural and neurochemical mechanisms. *Nat Hum Behav* 2:507–513.
- Brainard DH (1997) The psychophysics toolbox. *Spat Vis* 10:433–436.
- Censor N, Sagi D (2008) Benefits of efficient consolidation: short training enables long-term resistance to perceptual adaptation induced by intensive testing. *Vision Res* 48:970–977.
- Censor N, Karni A, Sagi D (2006) A link between perceptual learning, adaptation and sleep. *Vision Res* 46:4071–4074.
- Chalkia A, Van Oudenhove L, Beckers T (2020) Preventing the return of fear in humans using reconsolidation update mechanisms: a verification report of Schiller et al. (2010). *Cortex* 129:510–525.
- de Beukelaar TT, Woolley DG, Wenderoth N (2014) Gone for 60 seconds: reactivation length determines motor memory degradation during reconsolidation. *Cortex* 59:138–145.
- Fiorentini A, Berardi N (1981) Learning in grating waveform discrimination: specificity for orientation and spatial frequency. *Vision Res* 21:1149–1158.
- Furmanski CS, Engel SA (2000) Perceptual learning in object recognition: object specificity and size invariance. *Vision Res* 40:473–484.
- Grill-Spector K, Kushnir T, Hendler T, Malach R (2000) The dynamics of object-selective activation correlate with recognition performance in humans. *Nat Neurosci* 3:837–843.
- Hardwicke TE, Taqi M, Shanks DR (2016) Postretrieval new learning does not reliably induce human memory updating via reconsolidation. *Proc Natl Acad Sci USA* 113:5203–5211.
- Jeter PE, Doshier BA, Liu SH, Lu ZL (2010) Specificity of perceptual learning increases with increased training. *Vision Res* 50:1928–1940.
- Karni A, Sagi D (1993) The time course of learning a visual skill. *Nature* 365:250–252.
- Kleiner M, Brainard D, Pelli D, Ingling A, Murray R (2007) What's new in Psychtoolbox-3. *Perception* 36.
- Lee JLC, Nader K, Schiller D (2017) An update on memory reconsolidation updating. *Trends Cogn Sci* 21:531–545.
- Luyten L, Beckers T (2017) A preregistered, direct replication attempt of the retrieval-extinction effect in cued fear conditioning in rats. *Neurobiol Learn Mem* 144:208–215.
- Monfils MH, Cowansage KK, Klann E, Ledoux JE (2009) Extinction-Reconsolidation boundaries: key to persistent attenuation of fear memories. *Science* 324:951–955.
- Op de Beeck HP, Wagemans J, Vogels R (2007) Effects of perceptual learning in visual backward masking on the responses of macaque inferior temporal neurons. *Neuroscience* 145:775–789.
- Pelli DG (1997) The VideoToolbox software for visual psychophysics: transforming numbers into movies. *Spat Vis* 10:437–442.
- Sagi D (2011) Perceptual learning in vision research. *Vision Res* 51:1552–1566.
- Schiller D, Monfils MH, Raio CM, Johnson DC, Ledoux JE, Phelps EA (2010) Preventing the return of fear in humans using reconsolidation update mechanisms. *Nature* 463:49–53.
- Van Meel C, Daniels N, Op de Beeck HP, Baeck A (2016) Effect of tDCS on task relevant and irrelevant perceptual learning of complex objects. *J Vis* 16:13.
- Walker MP, Stickgold R (2016) Understanding the boundary conditions of memory reconsolidation. *Proc Natl Acad Sci USA* 113: E3991–E3992.
- Walker MP, Brakefield T, Morgan A, Hobson JA, Stickgold R (2002) Practice with sleep makes perfect: sleep-dependent motor skill learning. *Neuron* 35:205–211.
- Walker MP, Brakefield T, Hobson JA, Stickgold R (2003) Dissociable stages of human memory consolidation and reconsolidation. *Nature* 425:616–620.