Vision Research 227 (2025) 108543

Contents lists available at ScienceDirect

Vision Research

journal homepage: www.elsevier.com/locate/visres

Brief memory reactivation may not improve visual perception

Jun-Ping Zhu, Jun-Yun Zhang [*]

School of Psychological and Cognitive Sciences, and Beijing Key Laboratory of Behavior and Mental Health, Peking University, Beijing, China

ARTICLE INFO

Keywords:
Perceptual learning
Brief reactivation
Reconsolidation
Test–retest effect
Plasticity

ABSTRACT

Visual perceptual learning often requires a substantial number of trials to observe significant learning effects. Previously Amar-Halpert et al. (2017) have shown that brief reactivation (5 trials/day) is sufficient to improve the performance of the texture discrimination task (TDT), yielding comparable improvements to those achieved through full practice (252 trials/day). The finding is important since it would refine our understanding of learning mechanisms and applications. In the current study, we attempted to replicate these experiments using a larger number of observers and an improved experimental design. Using between-group comparison, we did find significant improvements in the reactivation group and the full-practice group as Amar-Halpert et al. (2017) showed. However, these improvements were comparable to those of the no-reactivation group with no exposure to the TDT task over the same period. Importantly, our within-group comparison showed that both the reactivation and no-reactivation groups exhibited additional significant improvements after further practicing the TDT task for an additional three days, demonstrating that the full-practice effect was significantly superior to the effects of brief memory reactivation or simple test–retest. Besides, when refining the constant stimuli method with fewer stimulus levels and more trials per level, we still observed comparable improvements brought by the reactivation and no-reactivation groups. Therefore, our results suggested that brief memory reactivation may not significantly contribute to the improvement of perceptual learning, and traditional perceptual training could still be a necessary and effective approach for substantial improvements.

1. Introduction

Visual perceptual learning refers to performance improvement on visual tasks through training (Lu & Dosher, 2022; Sagi, 2011; Watanabe & Sasaki, 2015). It has shown powerful real-world applications in improving the sensory performance of healthy individuals and rehabilitating clinical populations with various types of vision loss, such as amblyopia (Levi & Polat, 1996; Zhang et al., 2014), macular degeneration (Plank et al., 2014), and cortical blindness (Das et al., 2014; Herpich et al., 2019). However, a significant limitation of perceptual learning's practical applications is that it usually requires a long period of extensive practice for adequate performance enhancement (Jeter et al., 2010; Li et al., 2008). For example, a healthy adult's performance usually reaches a plateau after practicing for 5–10 daily sessions in a texture discrimination task (Karni & Sagi, 1991; Wang et al., 2013). Additionally, patients with vision impairments like cortical blindness, a form of vision loss caused by primary visual cortex damage, require months of daily practice to restore normal performance on a motion integration task in the blind field, making the training difficult to attain and sustain (Das et al., 2014; Huxlin et al., 2009).

Amar-Halpert et al. (2017) previously reported that brief reactivation of encoded visual memories was sufficient to improve visual perception. This study is grounded in the reactivation-reconsolidation framework, which claims that memories remain dynamic even after initial consolidation. Reactivation of memory through exposure to salient training stimuli can induce destabilization, triggering a reconsolidation process during which memories become susceptible to modification and can be enhanced or impaired (Lee et al., 2017). In the study of Amar-Halpert et al. (2017), observers in the reactivation group performed a texture discrimination task with 252 trials on day 1 to encode and consolidate memory. Subsequently, memory reactivation was conducted with only 5 trials for three consecutive days. The results showed that brief reactivations were sufficient to improve memory, as evidenced by the significant learning outcomes observed in the post-test on day 5, which were comparable to those of the full-practice group that performed the task with 252 trials per day. Besides, the reactivation group outperformed a no-progress control condition measuring two-session learning without memory reactivations (day 1 to day 2 in the

* Corresponding author.
E-mail address: zhangjunyun@pku.edu.cn (J.-Y. Zhang).

https://doi.org/10.1016/j.visres.2025.108543
Received 1 November 2023; Received in revised form 7 January 2025; Accepted 7 January 2025

full-practice group). They also established far-threshold reactivation and no-reactivation groups with pre-test and post-test spaced nine days apart and consistently found that the former outperformed the latter.

Several investigations using task interference have suggested the reactivation and reconsolidation process in perceptual learning (Bang et al., 2018; Dayan et al., 2016; Herszage & Censor, 2017; Huang et al., 2023; Shibata et al., 2017; Walker et al., 2003). For example, Bang et al. (2018) demonstrated that reconsolidation did occur after reactivation in visual perceptual learning. They asked observers to practice the detection of two orientations in a blockwise manner and found that the timing between the blocks (either short: 0 h or long: 3.5 h) led to either interference and performance decline, or no interference and performance improvement. These results suggested that reconsolidation occurred during the 3.5-hour interval following the reactivation of the trained orientation detection task. To the best of our knowledge, Amar-Halpert et al. (2017) is the first study reporting the reconsolidation phenomenon in the domain of visual perceptual learning. The finding of Amar-Halpert et al. (2017) is significant as it has challenged the fundamental principle of procedural learning theory, which states that practice makes perfect. Instead, their finding suggests a more efficient mechanism underlying improvement in visual perception, which has far-reaching clinical applications. The same research team has also generalized these results to other fields of procedural learning, including motor skill learning (Herszage et al., 2021), numeric domain (Schrift et al., 2022), and clinical populations like individuals with autism (Klorfeld-Auslender et al., 2022). This generalization across different memory fields and populations has great theoretical significance for the reconsolidation theory itself, given that the reconsolidation phenomenon has predominantly been based on Pavlovian fear-conditioning models in rodents since its initial discovery (Lee et al., 2017; Misanin et al., 1968; Nader et al., 2000; Schneider & Sherman, 1968).

More recently, Chen and de Beeck (2021) have investigated to what extent the similar effects of reactivation as Amar-Halpert et al. (2017) have shown in the texture discrimination task could be found in a paradigm that focuses on the learning of more complex visual objects. Chen and de Beeck (2021) conducted extensive measurements in more observers (N = 52). They found that although there was small progress in the reactivation group (25 trials/day), this improvement did not reach the same level of improvement as the group with full practice (800 trials/day), which was inconsistent with the result of the Amar-Halpert et al. (2017) study reporting comparable learning effects between reactivation and full practice. Previous evidence has shown that boundary conditions (such as memory type, memory strength, and age) are important determinants of whether memory is more or less susceptible to being reactivated and, possibly, disrupted (Auber et al., 2013). These boundary conditions under which memory does not undergo reconsolidation have been used to explain the contrasting results obtained across different studies. The limited effect of memory reactivation in Chen & de Beeck (2021) resonates with other domains in which it has proven hard to identify the boundary conditions under which a reactivation protocol is effective (Hardwicke et al., 2016). On the other hand, Chen and de Beeck (2021) explain the inconsistency with the positive findings of Amar-Halpert et al. (2017) as differences in the domain (texture vs. objects) and training protocols, considering that perceptual learning is sensitive to a lot of variables, including stimulus parameters.

Because of the theoretical and practical significance of brief reactivations in improving visual perceptual learning and the inconsistent results with different visual tasks, we opted to conduct an independent replication of the study by Amar-Halpert et al. (2017) and improve their experimental design. First, we adopted the same procedure as theirs in our reactivation and full-practice groups, but with a larger sample size (n = 23, as opposed to n = 12) to enhance statistical power. We established a no-reactivation group (n = 23) with no exposure to the TDT task over the same period as the other two groups, rather than utilizing the first two-session improvement of a full-practice group or a no-reactivation group (n = 7) with pre-test and post-test spaced nine days apart, which were implemented as the no-reactivation control condition in Amar-Halpert et al. (2017). Second, building upon the between-group design, we incorporated a within-group design in which both the reactivation and no-reactivation groups continued to practice the TDT task for an additional three days after the post-test on day 5. This extension allowed us to capture their full-practice progress from day 1 to day 8 and achieve within-group comparisons by comparing the full-practice effect with the reactivation-induced improvement in the reactivation group or single test–retest effect in the no-reactivation group. The within-group comparison can enhance the statistical power of the analysis by reducing variability and increasing the sensitivity to detect meaningful differences in the learning effects between reactivation and full practice (Kantowitz et al., 2009). Third, to weaken the inference about data overdispersion and yield more stable threshold estimates, a control experiment with a refined constant stimuli method by reducing the 14 SOA levels to 6 or 7 levels on new reactivation and new no-reactivation groups was conducted. Finally, as the thresholds were estimated by fitting the psychometric curves and different fitting methods might induce varying threshold estimates (Kingdom & Prins, 2010; Manning et al., 2018), to see whether the results were affected by fitting methods, we estimated thresholds using two approaches: the same fitting method as Amar-Halpert et al. (2017) and the psignifit4 fitting method which allows accurate Bayesian estimation of psychometric functions for (potentially) overdispersed data (Schütt et al., 2016).

2. Methods

2.1. Observers and apparatus

Eighty-five observers (aged 21.9 ± 3.3 years) with normal or corrected-to-normal vision participated in this study: 23 in the reactivation group, 23 in the no-reactivation group, 23 in the full-practice group, 8 in the new reactivation group, and 8 in the new no-reactivation group. All were new to psychophysical experiments and were unaware of the purposes of the study. The study was conducted in accordance with the Declaration of Helsinki and was approved by the Peking University Institution Review Board. Informed consent was obtained from each observer before data collection.

The stimuli were generated with Psychtoolbox-3 software (Kleiner et al., 2007; Pelli, 1997) and presented on a 21-inch Sony G520 color monitor (1024-pixel × 768-pixel resolution, 0.39-mm × 0.39-mm pixel size, 75-Hz frame rate). The minimal and maximal luminance of the monitor was 0.3 and 74.3 cd/m^2, respectively. A chin-and-head rest helped stabilize the observer's head. The viewing was binocular at a distance of 105 cm. Experiments were run in a dimly lit room. Responses were collected via the computer keyboard.

2.2. Stimuli and procedure

Stimuli. The texture discrimination task (TDT) was nearly identical to the one in Amar-Halpert et al. (2017) study. It primarily included a target frame and a mask (Fig. 1a). The target frame occupied an area of 15.1° × 15.1° at a viewing distance of 105 cm. It consisted of a 19 × 19 array of white bars displayed on a black screen background. Within this array, a target was presented at the lower right visual quadrant, centered at 5.7° from the texture center with the target center location jittering ± 1.3° from trial to trial. The target configuration defined by three diagonal bars fixed at 45° was vertical or horizontal. The target configuration was embedded in a background of horizontal bars that were 0.46° × 0.03° each and spaced 0.79° apart in the array. The position of each bar was slightly jittered from trial to trial, ranging from 0° to 0.11°. In the center of the array, a randomly oriented letter T or L (0.48° × 0.37°) was presented to control fixation. The mask was a same-sized field consisting of randomly oriented V-shaped patterns and a central compound pattern of superimposed T and L.

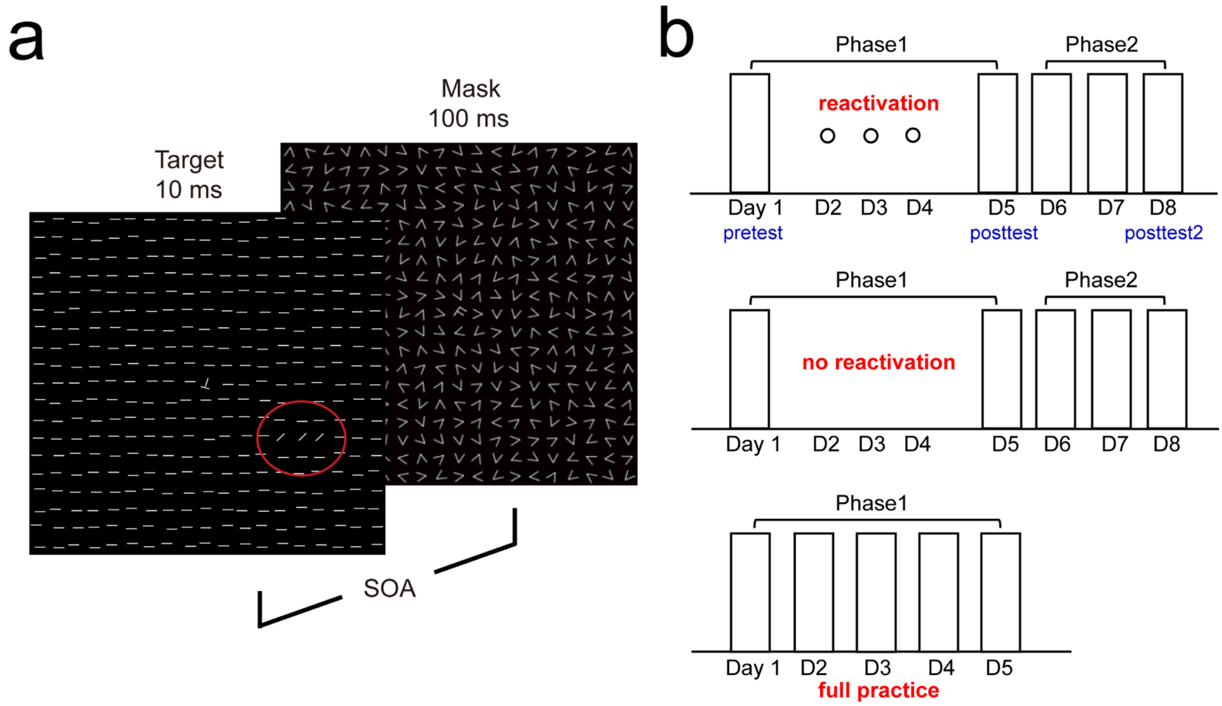

Fig. 1. Stimuli and study design **a.** Stimulus configuration of the texture discrimination task (TDT). The red circle was not present in the actual stimuli. **b.** The study design included a between-group design and a within-group design (main experiment). The between-group design consisted of the reactivation group, the no-reactivation group, and the full-practice group. The within-group design in the reactivation and the no-reactivation groups covered Phase 1 and Phase 2. The pre-test, post-test, and post-test2 thresholds were measured on day 1, day 5, and day 8 respectively. In the control experiment, the new reactivation and no-reactivation groups only experienced Phase 1.

Procedure. A standard trial of the texture discrimination task in the current study was nearly identical to that of Amar-Halpert et al. (2017). Each trial began with a 400 ms presentation of a fixation cross, followed by a 500 ms blank screen display. Then a target frame was briefly presented for 13.3 ms, followed, at various stimulus onset asynchronies (SOAs, measured from the onset of the target to the onset of the mask), by a 100 ms patterned mask. After the mask, the screen went blank until the observer made a response. The observers were asked to make two responses: first to report the foveal letter (T or L), and then to report the orientation of the target configuration (horizontal or vertical). Immediate auditory feedback was provided only for the incorrect foveal letter identification. There was a 250 ms inter-trial interval. The average accuracy of the foveal letter identification task is approximately 95 %, indicating effective foveal fixation.

The study employed the constant stimuli method, in which several predetermined stimulus levels were used and each level consisted of a fixed number of trials. The stimulus levels were stimulus onset asynchronies (SOAs, measured from the onset of the target to the onset of the mask), which were multiples of 13.3 ms frame duration, ranging from 3 frames to 26 frames (40, 67, 80, 107, 120, 147, 160, 187, 200, 227, 240, 267, 307, 347 ms). In the main experiment, the 14 SOA levels were randomized across all trials, with 18 trials per SOA level. Thus, a standard block comprised a total of 252 trials. In the control experiment, the constant stimuli method was refined by reducing the 14 SOA levels to 6 or 7 levels, with each level containing 42 or 36 trials, maintaining the total number of trials at 252.

Before the formal experiment, a pretraining block consisting of 10 trials at a 347 ms SOA was administered repeatedly until observers reached a 90 % accuracy rate. Observers who completed ten pretraining blocks without attaining the required accuracy were excluded from the study. Those who successfully passed the pretraining phase then progressed to a familiarization block comprising 14 trials, each corresponding to one of the 14 SOAs. Subsequently, the formal experiment commenced.

2.3. Experimental design

The main experiment included three groups of observers: the reactivation group, the no-reactivation group, and the full-practice group (Fig. 1b). In the first phase, we attempted to replicate the procedure in the Amar-Halpert et al. (2017) study. All observers performed the TDT with 252 trials on day 1 and day 5 to measure the pre-test and post-test threshold respectively. The operation in the middle of three days divided observers into three groups. From day 2 to day 4, the full-practice group still practiced the TDT with 252 trials per day, and the brief-reactivation group practiced the TDT with only 5 trials per day to reactivate memory while no operation was carried out on the no-reactivation group. Reactivation trials were set individually at one of the 14 SOAs that was closest to each observer's pre-test threshold (See Table S1 for a specific value of each observer). The day numbers (e.g., 1–5) typically, but not always, represent consecutive days. Most observers finished Phase 1 within 6 days (23/23, 22/23, and 21/23 in the reactivation, no-reactivation, and full-practice groups, respectively). In the second phase, the reactivation and no-reactivation groups continued to practice the TDT with 252 trials per day for three days (from day 6 to day 8). This operation allowed us to obtain their full-practice performance (from day 1 to day 8) and compare it with their previous reactivation-induced improvement or single test–retest effect (from day 1 to day 5). These within-group comparisons, focusing on individual changes within the same group, help control for individual differences, thereby providing more reliable results than between-group comparisons. Additionally, the within-group comparison can enhance the statistical power of the analysis by reducing variability and increasing the sensitivity to detect meaningful differences (Kantowitz et al., 2009).

In the control experiment, the constant stimuli method was refined by reducing the 14 SOA levels to 6 or 7 levels, with each level containing 42 or 36 trials, maintaining the total number of trials at 252. Sixteen new observers were randomly assigned to the new reactivation group (n = 8) or the new no-reactivation group (n = 8). All observers in the new reactivation group and half of the observers in the new no-reactivation group experienced 6 SOA levels (40, 67, 107, 160, 240, 347 ms), with 42 trials per SOA level. The other half of observers in the new no-reactivation group underwent 7 SOA levels (40, 80, 120, 160, 200, 267, 347 ms), with 36 trials per SOA level. Reactivation trials were set individually at one of the 6 ~ 7 SOAs that was closest to each observer's pre-test threshold (See Table S1 for a specific value of each observer). Using the modified constant stimuli method, the new reactivation and no-reactivation groups only experienced Phase 1.

2.4. Data fitting and statistical analysis

To evaluate the impact of different data fitting methods on the results, we employed two fitting methods to fit psychometric curves for threshold estimates.

The first fitting method is consistent with the approach described by Amar-Halpert et al. (2017). The threshold was calculated for each standard block (252 trials) using the Weibull fit for the psychometric curve, with slope β and finger error (mistyping) parameter 1 – p, yielding the function:

$$P(t) = p\{1-\frac{1}{2}\exp[-\left(\frac{t}{T}\right)^{\beta}]\}+\frac{1-p}{2} = \frac{1}{2}\{1+p[1-\exp[-\left(\frac{t}{T}\right)^{\beta}]]\}$$

where P(t) is the measured probability of correct response; t represents the SOA levels; finger error parameter, which takes stimulus-independent errors (e.g., attention lapses, response-key confusion) into account, is a free parameter within a range ($0 < 1 - p < 1$); T is the estimated threshold for each curve, defined as the SOA for which 81.6 % of responses were correct when $p = 1$. Weibull fit was computed using a maximum likelihood method, assuming a binomial process (Wichmann & Hill, 2001).

The second method is using the psignifit4 software package (see http://bootstrap-software.org/psignifit/) to fit psychometric curves and estimate thresholds (Schütt et al., 2016). Here the psychometric function modeling was extended from the standard binomial to a beta-binomial model to enable accurate Bayesian estimation of psychometric functions even for overdispersed data. Psychometric curves for each observer were generated by fitting the data with a Weibull function. Using the chosen sigmoid family S(x; m, w), the psychometric function ψ is defined with two additional parameters λ and γ for the upper and lower asymptote, scaling the sigmoid function:

$$\psi(x;\, m,\, w,\, \lambda,\, \gamma) = \gamma + (1-\lambda-\gamma)S(x;\, m,\, w)$$

where threshold m is the stimulus level at which 81.6 % of responses were correct when $\lambda = 0$ (to maintain consistency with the first fitting method); w represents the width (difference between the stimulus levels for which the unscaled function reaches 0.05 and 0.95 respectively); λ represents the lapse rate (the difference between the upper asymptote and 1); γ represents the guess rate (the difference between the lower asymptote and 0). γ is fixed at 0.5 and the lapse rate λ is free.

For both fitting methods, to evaluate how well the psychometric curves capture the empirical data of each individual, we assess goodness-of-fit by calculating deviance which is recommended for binomial data (Wichmann & Hill, 2001):

$$D = 2\sum_{i=1}^{K}\left\{n_i y_i \log\left(\frac{y_i}{p_i}\right) + n_i(1-y_i)\log\left(\frac{1-y_i}{1-p_i}\right)\right\}$$

where K denotes the number of SOA levels, n_i: the number of trials in SOA level i,y_i: the observer's response accuracy in SOA level i, p_i: the response accuracy predicted by the fitted model.

For correct models, deviance for binomial data was asymptotically distributed as χ^2_k, where K denoted the number of SOA levels and a χ^2 probability of < 0.05 is considered to indicate a poor fit of the model (Hietanen et al., 2022; Lasagna et al., 2020; Wichmann & Hill, 2001). In our data, both fitting methods indicated the same observer (R23) with poor goodness-of-fit on day 1 (see Figs. S1 for detailed information). As the statistical analyses produced similar results regardless of whether this observer was included or not in the analyses, we decided to keep this observer in the threshold analyses.

The learning effect was evaluated by the improvement in thresholds. Individual threshold improvement from the pre-test to the post-test was calculated as 100 %× ($Threshold_{pretest}$ – $Threshold_{posttest}$)/$Threshold_{pretest}$ and then averaged across observers to obtain the mean percent improvement (MPI). To evaluate the progress following Phase2 training, individual further threshold improvement from post-test to post-test2 was calculated as 100 % × ($Threshold_{post\text{-}test}$ – $Threshold_{post\text{-}test2}$)/$Threshold_{post\text{-}test}$, and individual total threshold improvement from pre-test to post-test2 was calculated as 100 % ×($Threshold_{pretest}$ – $Threshold_{posttest2}$)/$Threshold_{pretest}$.

All analyses were conducted using open-source JASP software version 0.17.2.1 (Wagenmakers et al., 2018). Improvements in SOA thresholds were compared against the value of 0 through a one-sample *t*-test. Within-group comparisons were performed using a paired samples *t*-test or one-way repeated measures analysis of variance (ANOVA). Comparisons between the two groups were conducted using both classical and Bayesian independent samples t-tests.

3. Results

3.1. Main experiment_Phase1: Comparing learning effects of reactivation with no reactivation and full practice

In Phase 1 covering five days, we attempted to replicate the experiments of Amar-Halpert et al. (2017) with a larger number of observers. A total of 69 observers were randomly assigned into three groups, with 23 observers in each group (Fig. 1b). To assess the impact of different fitting methods on the results, we estimated thresholds using two approaches: the same fitting method as Amar-Halpert et al. (2017) (see Figs. S1-1, S2-1, S3-1 for individuals' data fitting) and the psignifit4 fitting method (see Figs. S1-2, S2-2, S3-2 for individuals' data fitting). The results of the same fitting method were presented unless specified.

In the reactivation group, thresholds in the post-test (day 5) were significantly reduced compared to those in the pre-test (day 1) (Fig. 2a (i), $mean_{pretest} = 127.3 \pm 10.2$ ms, $mean_{posttest} = 85.8 \pm 4.6$ ms, $t_{22} = 3.99$, $p < 0.001$, Cohen's d = 0.83; psignifit4: Fig. 2c(i), $mean_{pretest} = 143.1 \pm 14.4$ ms, $mean_{posttest} = 88.9 \pm 4.8$ ms, $t_{22} = 3.92$, $p < 0.001$, Cohen's d = 0.82). The TDT performance improved significantly (Fig. 2b, MPI = 27.1 ± 4.4 %, $t_{22} = 5.93$, $p < 0.001$, Cohen's d = 1.24; psignifit4: Fig. 2d, MPI = 28.0 ± 5.4 %, $t_{22} = 5.16$, $p < 0.001$, Cohen's d = 1.08). The mean percent improvement of the reactivation group in Amar-Halpert et al. (2017) was also significant (MPI = 20.6 ± 5.5 %).

In the no-reactivation group, thresholds in the post-test (day 5) were significantly lower than those in the pre-test (day 1) (Fig. 2a(ii), $mean_{pretest} = 117.9 \pm 6.0$ ms, $mean_{posttest} = 90.2 \pm 6.2$ ms, $t_{22} = 5.46$, $p < 0.001$, Cohen's d = 1.14; psignifit4: Fig. 2c(ii), $mean_{pretest} = 128.4 \pm 9.7$ ms, $mean_{posttest} = 92.6 \pm 6.5$ ms, $t_{22} = 3.9$, $p < 0.001$, Cohen's d = 0.82), with a threshold decrease of 27.7 ± 5.1 ms (psignifit4: 35.8 ± 9.1 ms) from pre-test to post-test. The TDT performance improved significantly (Fig. 2b, MPI = 22.7 ± 3.8 %, $t_{22} = 5.91$, $p < 0.001$, Cohen's d = 1.23; psignifit4: Fig. 2d, MPI = 23.9 ± 4.5 %, $t_{22} = 5.32$, $p < 0.001$, Cohen's d = 1.11). However, the no-reactivation group (n = 7) in Amar-Halpert et al. (2017) showed little improvement, with a threshold decrease of 7.6 ± 3.3 ms from the pre-test to the post-test.

In the full-practice group, thresholds in the post-test (day 5) were significantly lower than those in the pre-test (day 1) (Fig. 2a(iii), $mean_{pretest} = 125.0 \pm 8.7$ ms, $mean_{posttest} = 88.0 \pm 5.3$ ms, $t_{22} = 5.8$, $p < 0.001$, Cohen's d = 1.21; psignifit4: Fig. 2c(iii), $mean_{pretest} = 135.4 \pm 12.1$ ms, $mean_{posttest} = 89.4 \pm 6.3$ ms, $t_{22} = 5.39$, $p < 0.001$, Cohen's d = 1.1). The TDT performance improved significantly (Fig. 2b, MPI = 27.3 ± 3.4 %, $t_{22} = 8.04$, $p < 0.001$, Cohen's d = 1.68; psignifit4: Fig. 2d,

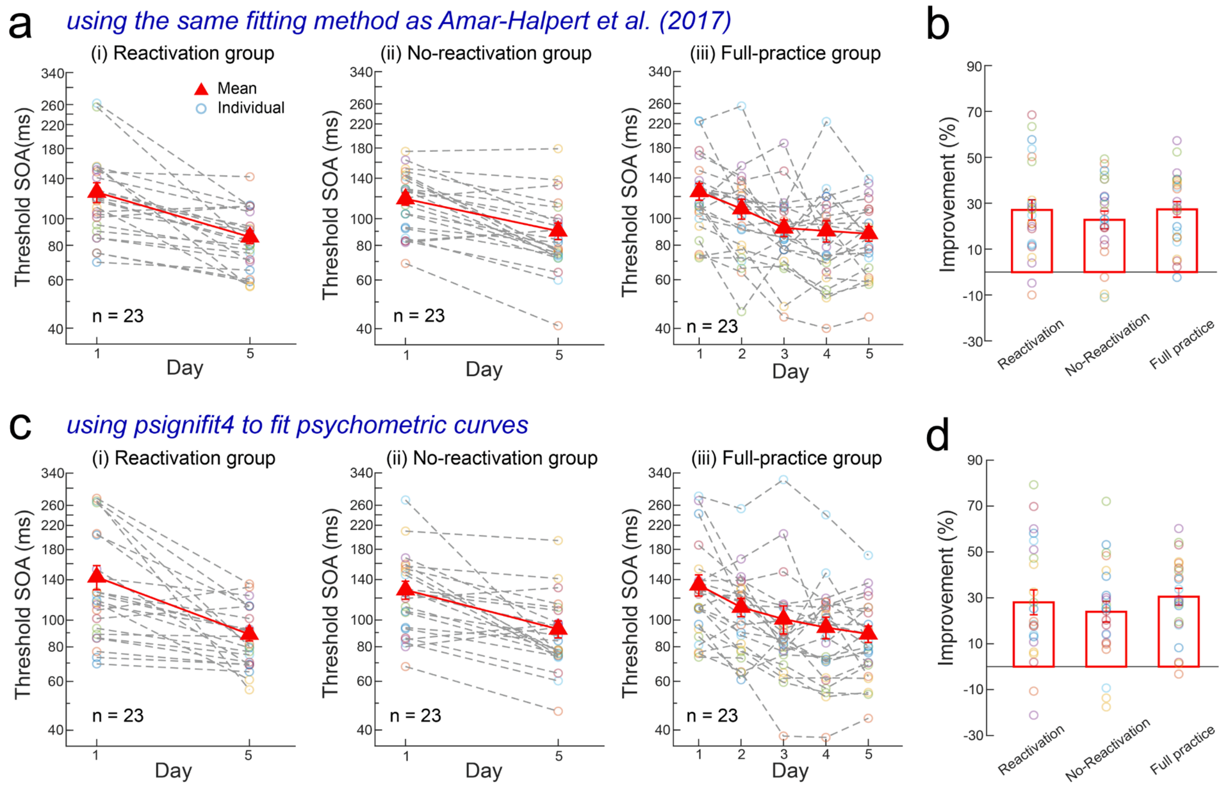

Fig. 2. Phase 1: Improvements in TDT performance from pre-test (day 1) to post-test (day 5) for the three groups under two fitting methods. **a & c.** The thresholds changed as days in the reactivation group (i), the no-reactivation group (ii), and the full-practice group (iii) under the same fitting method as Amar-Halpert et al. (2017) (a) and the psignifit4 fitting method from Schütt et al., 2016 (c). Solid triangles and hollow circles represented mean and individual values, respectively. The threshold SOA of the y-axis was logarithmic. **b & d.** Percent improvements in TDT performance from pre-test to post-test for the three groups under the same fitting method (b) and the psignifit4 fitting method (d). Circles represented the individuals' data. Error bars indicated ± 1 standard error of the mean.

MPI = 30.5 ± 3.6 %, t_{22} = 8.36, p < 0.001, Cohen's d = 1.74). The learning progress from day 1 to day 2 in our full-practice group was also significant (MPI = 12.5 ± 4.2 %, t_{22} = 2.96, p = 0.007, Cohen's d = 0.62; psignifit4: MPI = 12.4 ± 4.0 %, t_{22} = 3.09, p = 0.005, Cohen's d = 0.65). The total learning effect of the full-practice group in Amar-Halpert et al. (2017) was also significant (MPI = 26.6 ± 5.9 %), but their day1-to-day2 improvement was insignificant (MPI = 2.9 ± 5.8 %).

A classical independent sample t-test revealed no significant difference in learning improvements between the reactivation group and the no-reactivation group (t_{44} = 0.75, p = 0.46, Cohen's d = 0.22; psignifit4: t_{44} = 0.58, p = 0.57, Cohen's d = 0.17). A Bayesian independent-sample t-test also supports the null hypothesis, with a Bayes factor (BF_{10}) of 0.37 (psignifit4: BF_{10} = 0.34), representing the ratio of the likelihood of the observed data under the alternative hypothesis to the likelihood under the null hypothesis. A BF_{10} of 0.37 or 0.34 indicated anecdotal evidence in favor of the null hypothesis, according to the interpretation of the Bayes factor magnitude (Johnson et al., 2022; Wagenmakers et al., 2018). The finding suggested that brief reactivations did not yield additional gains in learning improvement compared to a no-reactivation control condition, with these improvements likely attributable to a significant retest effect. Besides, both classical and Bayesian independent samples t-tests indicated no significant difference in learning improvements between the reactivation group and the full-practice group (t_{44} = 0.03, p = 0.98, Cohen's d = 0.01; BF_{10} = 0.29, moderate evidence for the null hypothesis; psignifit4: t_{44} = 0.38, p = 0.71, Cohen's d = 0.11; BF_{10} = 0.31, moderate evidence for the null hypothesis).

In summary, the results from both fitting methods consistently demonstrated that the significant learning effect in the reactivation group was comparable to that in the no-reactivation group, and more likely to reflect a significant retest effect or fast learning occurring in session.

3.2. Main experiment_Phase2: Continued training in the reactivation and no-reactivation groups

The results in Phase 1 for the three groups and the results in the study of Amar-Halpert et al. (2017) were derived from between-group comparisons, which could be affected by individual differences. And it was still in doubt whether there was potential for further progress for the significant retest effects in the reactivation and no-reactivation groups. Therefore, after the post-test on day 5, all 23 observers in the reactivation group and 20 observers in the no-reactivation group continued to practice the TDT for an additional three days (Phase 2, spanning from day 6 to day 8) with 252 trials per day. This extended practice aimed to capture their full-practice improvements over two phases (from day 1 to day 8) for within-group comparison with Phase 1 improvements (from day 1 to day 5).

In the reactivation group, thresholds in the post-test2 were significantly lower than those in the post-test (Fig. 3a, mean$_{\text{post-test}}$ = 85.8 ± 4.6 ms, mean$_{\text{post-test2}}$ = 68.4 ± 4.1 ms, t_{22} = 3.69, p = 0.001, Cohen's d = 0.77). The TDT performance improved significantly in Phase2 (Fig. 3b (i), MPI$_{\text{Phase2}}$ = 17.8 ± 5.1 %, t_{22} = 3.50, p = 0.002, Cohen's d = 0.73; psignifit4: Fig. 3b(ii), MPI$_{\text{Phase2}}$ = 19.8 ± 4.9 %, t_{22} = 4.05, p < 0.001, Cohen's d = 0.85). The full-practice improvements over the two phases were significantly greater than the Phase 1 improvements (Fig. 3b(i), MPI$_{\text{Phase1}}$ = 27.1 ± 4.4 %, MPI$_{\text{total}}$ = 40.7 ± 5.0 %, t_{22} = 3.67, p = 0.001, Cohen's d = 0.77; psignifit4: Fig. 3b(ii), MPI$_{\text{Phase1}}$ = 28.0 ± 5.4 %, MPI$_{\text{total}}$ = 43.2 ± 5.3 %, t_{22} = 3.78, p = 0.001, Cohen's d = 0.79). Notably, both classical and Bayesian independent samples t-tests indicated that the reactivation group's full-practice improvements were also greater than those of the full-practice group (t_{44} = 2.21, p = 0.03, Cohen's d = 0.65; BF_{10} = 2.02, anecdotal evidence for the alternative hypothesis; psignifit4: t_{44} = 1.96, p = 0.057, Cohen's d = 0.58; BF_{10} =

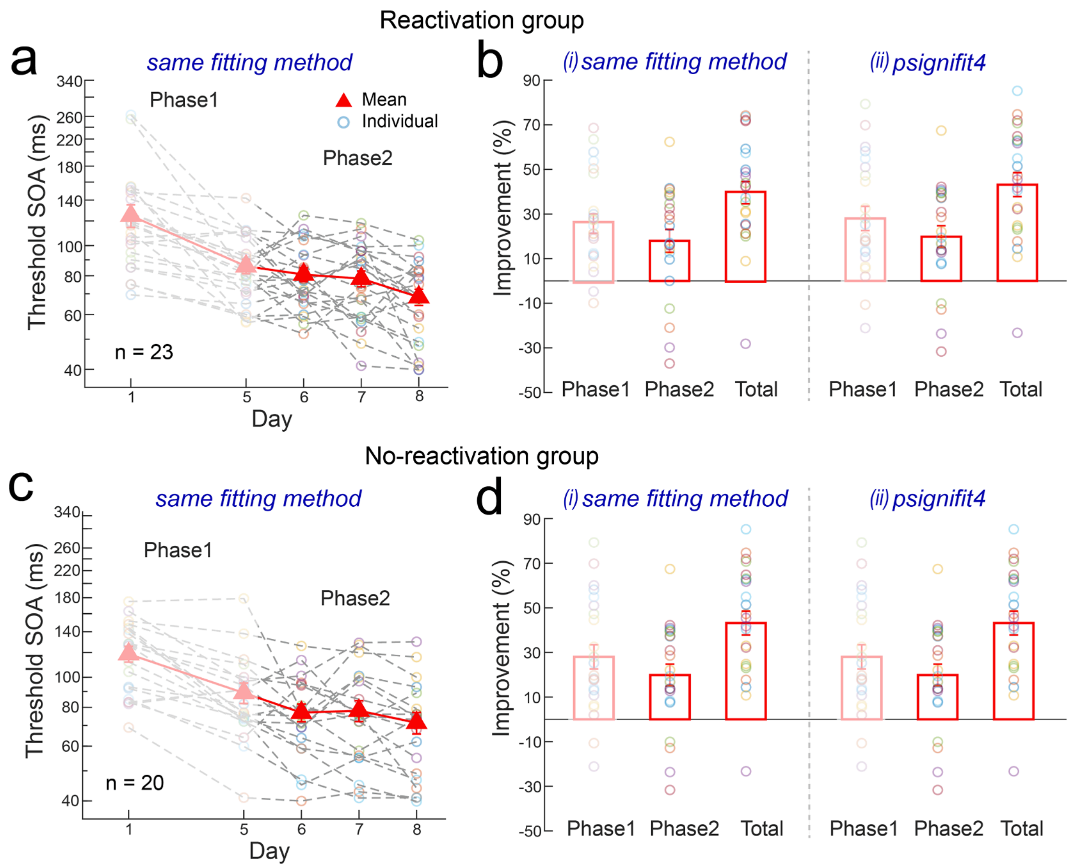

Fig. 3. Phase 2: Continued training of TDT in the reactivation and no-reactivation groups under two fitting methods. **a & c**. The thresholds in Phase 1 (shaded) and Phase 2 (unshaded) in the reactivation group (a) and the no-reactivation group (c) under the same fitting method as Amar-Halpert et al. (2017). Solid triangles and hollow circles represented mean and individual values, respectively. The threshold of the coordinate axis was logarithmic. **b & d.** Percent improvements in TDT performance in Phase 1, Phase 2, and two phases (total) under the same fitting method as Amar-Halpert et al. (2017) (i) and the second fitting method from Schütt et al., 2016 (ii) in the reactivation group (b) and the no-reactivation group (d). Circles represented the individuals' data. Error bars indicated ± 1 standard error of the mean.

1.34, anecdotal evidence for the alternative hypothesis).

Similarly, in the no-reactivation group, thresholds in post-test2 were also significantly reduced compared to those in the post-test (Fig. 3c, mean$_{\text{post-test}}$ = 89.0 ± 6.7 ms, mean$_{\text{post-test2}}$ = 71.3 ± 5.7 ms, t_{19} = 2.93, p = 0.009, Cohen's d = 0.66), and the TDT performance improved significantly during Phase2 (Fig. 3d(i), MPI$_{\text{Phase2}}$ = 17.0 ± 5.6 %, t_{19} = 3.01, p = 0.007, Cohen's d = 0.67; psignifit4: Fig. 3d(ii), MPI$_{\text{Phase2}}$ = 19.1 ± 5.7 %, t_{19} = 3.37, p = 0.003, Cohen's d = 0.75). The full-practice improvements over the two phases were significantly greater than the Phase 1 improvements (Fig. 3d(i), MPI$_{\text{Phase1}}$ = 24.1 ± 4.0 %, MPI$_{\text{total}}$ = 38.1 ± 4.3 %, t_{19} = 3.23, p = 0.004, Cohen's d = 0.72; psignifit4: Fig. 3d (ii), MPI$_{\text{Phase1}}$ = 25.1 ± 4.8 %, MPI$_{\text{total}}$ = 40.0 ± 4.8 %, t_{19} = 3.40, p = 0.003, Cohen's d = 0.76). Notably, both classical and Bayesian independent samples t-tests showed that the no-reactivation group's full-practice improvements were also greater than those of the full-practice group under the same fitting method (t_{41} = 2.01, p = 0.05, Cohen's d = 0.61; BF_{10} = 1.46, anecdotal evidence for the alternative hypothesis), although this difference was not observed using the psignifit4 fitting method (t_{41} = 1.60, p = 0.12, Cohen's d = 0.49; BF_{10} = 0.83, anecdotal evidence for the null hypothesis).

In summary, both the reactivation and no-reactivation groups exhibited significant further improvements during the continued training phase, suggesting that the gains in Phase 1 are significant but may be insufficient to improve the performance of the texture discrimination task. Besides, our results also indicated that spaced full practice over about 8 days in the reactivation and no-reactivation groups may be better than mass training over about 5 days in the full-practice group.

3.3. Control experiment: Comparing learning improvements between the reactivation and no-reactivation groups using a modified constant stimuli method

The main experiment adopted the constant stimuli method in Amar-Halpert et al. (2017) including 14 SOA levels, each comprising 18 trials. This methodology may feature an excessive number of SOA levels and inadequate trials per level, which could lead to non-stationarity in observer behavior. According to Blackwell (1952), it is preferable to measure more trials per level rather than more levels, as small blocks of trials weaken the inference about overdispersion and yield more stable threshold estimates. Therefore, we refined the constant stimuli method by reducing the 14 SOA levels to 6 or 7 levels, with each level containing 42 or 36 trials, maintaining the total number of trials at 252. Sixteen new observers were randomly divided into two groups: the new reactivation group (n = 8) and the new no-reactivation group (n = 8).

In the new reactivation group, thresholds in the post-test (day 5) were significantly reduced compared to those in the pre-test (day 1) (Fig. 4a(i), mean$_{\text{pretest}}$ = 129.1 ± 11.4 ms, mean$_{\text{post-test}}$ = 96.3 ± 10.2 ms, t_7 = 4.92, p = 0.002, Cohen's d = 1.74; psignifit4: Fig. 4c(i), mean$_{\text{pretest}}$ = 136.3 ± 16.7 ms, mean$_{\text{post-test}}$ = 95.7 ± 10.1 ms, t_7 = 3.12, p = 0.02, Cohen's d = 1.10). The TDT performance improved significantly (Fig. 4b, MPI = 25.2 ± 5.2 %, t_7 = 4.88, p = 0.002, Cohen's d = 1.72; psignifit4: Fig. 4d, MPI = 27.2 ± 6.9 %, t_7 = 3.92, p = 0.006, Cohen's d = 1.39).

Similarly, in the new no-reactivation group, thresholds in the post-test (day 5) were significantly lower than those in the pre-test (day 1)

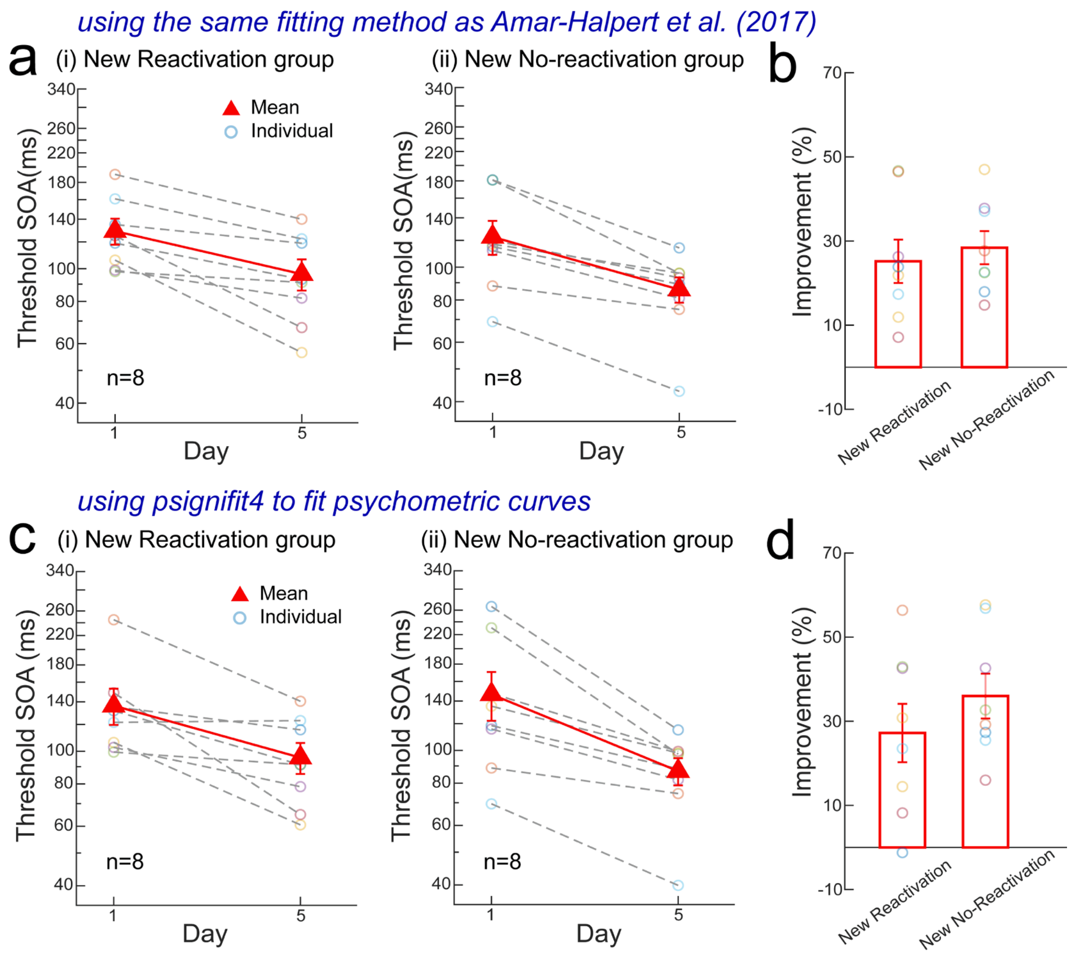

Fig. 4. Perceptual learning of TDT from pre-test (day 1) to post-test (day 5) in the new reactivation and no-reactivation groups using the modified constant stimuli method. **a & c**. The thresholds changed as days in the new reactivation group (i) and the new no-reactivation group (ii) under the same fitting method as Amar-Halpert et al. (2017) (a) and the psignifit4 fitting method from Schütt et al., 2016 (c). Solid triangles and hollow circles represented mean and individual values, respectively. The threshold SOA of the y-axis was logarithmic. **b & d.** Percent improvements in TDT performance from pre-test to post-test for the two groups under the same fitting method (b) and the psignifit4 fitting method (d). Circles represented the individuals' data. Error bars indicated ± 1 standard error of the mean.

(Fig. 4a(ii), mean$_{\text{pretest}}$ = 122.9 ± 14.1 ms, mean$_{\text{post-test}}$ = 85.9 ± 7.4 ms, t_7 = 4.16, p = 0.004, Cohen's d = 1.47; psignifit4: Fig. 4c(ii), mean$_{\text{pretest}}$ = 146.5 ± 24.1 ms, mean$_{\text{post-test}}$ = 86.8 ± 8.0 ms, $t_7 = 3.24$, p = 0.01, Cohen's d = 1.15). The TDT performance also improved significantly (Fig. 4d, MPI = 28.4 ± 3.9 %, $t_7 = 7.20$, $p < 0.001$, Cohen's d = 2.54; psignifit4: Fig. 4d, MPI = 36.0 ± 5.3 %, $t_7 = 6.75$, $p < 0.001$, Cohen's d = 2.39).

Both classical and Bayesian independent samples t-tests indicated no significant difference in learning improvements between the two groups ($t_{14} = 0.50$, $p = 0.63$, Cohen's d = 0.25; $BF_{10} = 0.47$, anecdotal evidence for the null hypothesis; psignifit4: $t_{14} = 1.0$, $p = 0.33$, Cohen's d = 0.50; $BF_{10} = 0.60$, anecdotal evidence for the null hypothesis). Therefore, the results of the control experiment further indicated that reactivation did not yield additional gains in learning improvement compared to a no-reactivation condition, which is consistent with the results of the main experiment.

3.4. Finger error/Lapse rate and goodness-of-fit under two fitting methods

Finger errors or lapse rates reflected stimulus-independent errors (e.g., attention lapses, response-key confusion). Their values in each fitting method were reported in Fig. 5 (also see Supplementary Tables S2 ~ S6). One can readily see that finger error/lapse rate values showed a trend of decrease across sessions (along with a decrease in thresholds), suggesting increasingly reliable judgments as training progressed.

Statistical analyses of the finger errors/lapse rate values across days were conducted for each group. When using the same fitting method as Amar-Halpert et al. (2017), a one-way repeated measures ANOVA showed that finger error values showed a significant decrease from day 1 (pre-test) to the last three days (reactivation group: Fig. 5a(i), $ps < 0.02$; full-practice group: Fig. 5a(iii), $ps < 0.04$) and a significant reduction from day 1 to the other four days (no-reactivation group: Fig. 5a(ii), $ps < 0.001$). Paired samples t-tests indicated that finger error values decreased significantly from day 1 to day 5 for both the new reactivation and new no-reactivation groups (Fig. 5a(iv), $ps < 0.01$).

Similarly, when using the psignifit4 fitting method, a one-way repeated measures ANOVA showed that the lapse rate values in the reactivation group showed no significant main effect of days (Fig. 5b(i), $p = 0.06$), but a paired samples *t*-test showed a significant decrease from day 1 to day 8 ($p = 0.04$). In the no-reactivation group, lapse rate values decreased significantly from day 1 to the other four days (Fig. 5b(ii), $ps < 0.03$). In the full-practice group, lapse rate values exhibited a significant reduction from the first two days to the last day (Fig. 5b(iii), $ps < 0.03$). Additionally, the lapse rate values of the new reactivation and new no-reactivation groups showed no significant changes from day 1 to day 5 (Fig. 5b(iv), $ps > 0.1$).

To evaluate how well the psychometric curves capture the empirical data of each individual, we assess goodness-of-fit by calculating deviance (Wichmann & Hill, 2001), in which smaller deviance indicates better goodness-of-fit (Haynes et al., 2024; Su et al., 2024). In particular,

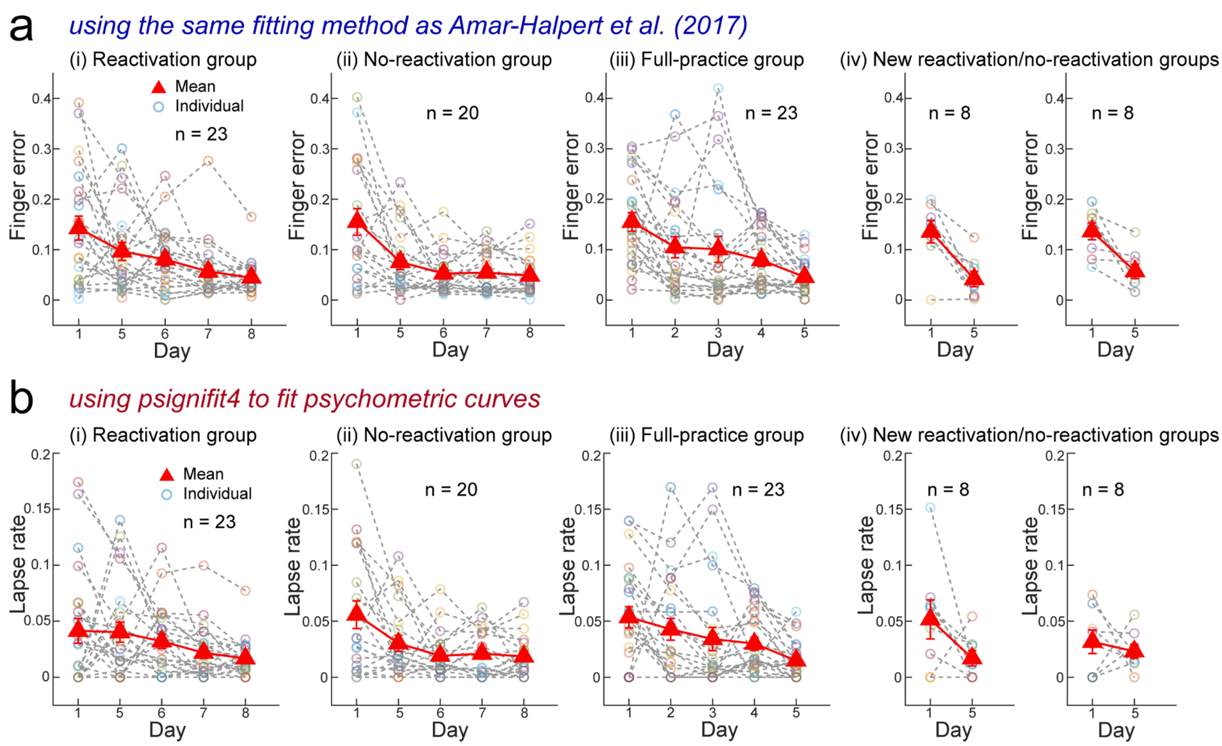

Fig. 5. Finger error/lapse rate across days in each group under two fitting methods. **a & b.** Finger error in the same fitting method as Amar-Halpert et al. (2017) (a) or lapse rate in the psignifit4 fitting method (b) changed as days in the reactivation group (i), the no-reactivation group (ii), the full-practice group (iii), the new reactivation group, and the new no-reactivation group (iv). Solid triangles and hollow circles represented mean and individual values, respectively. Error bars indicated ± 1 standard error of the mean. Note: 20 observers in the no-reactivation group finished the two phases of the operation.

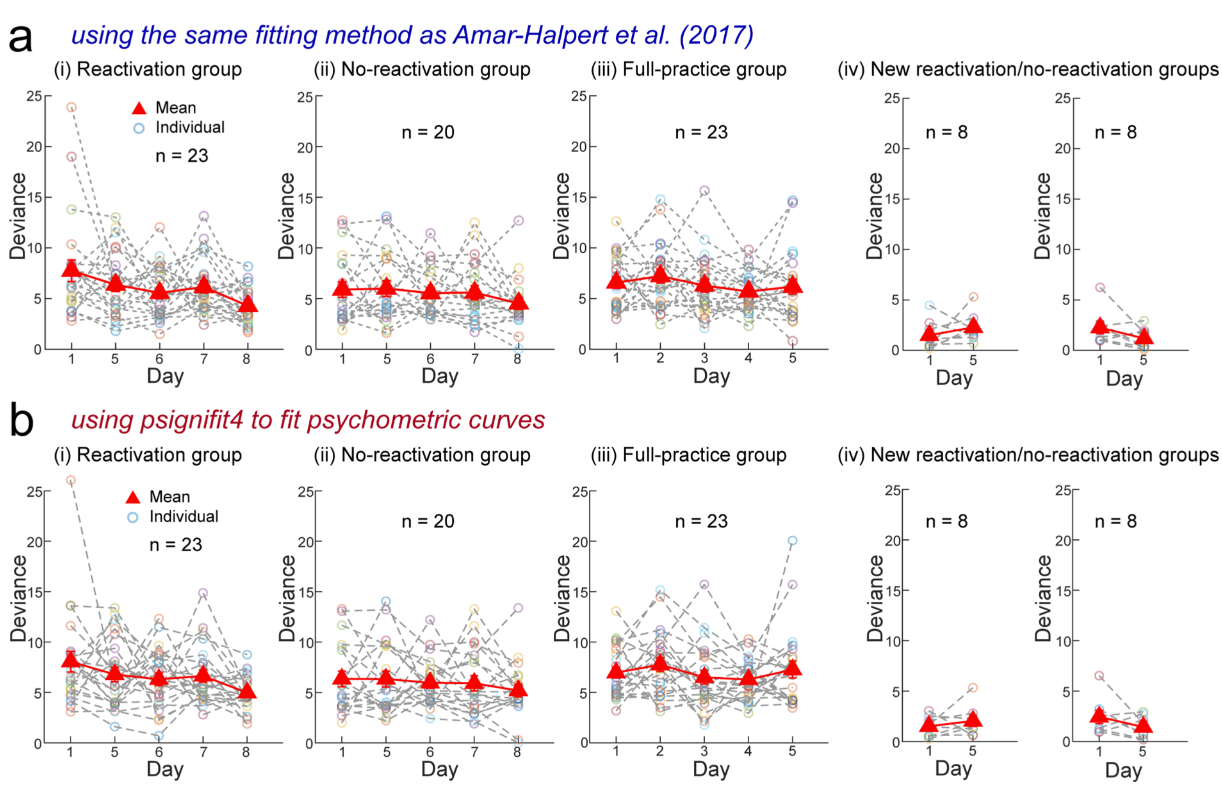

Fig. 6. Goodness-of-fit (deviance values) across days for each group under two fitting methods. **a & b.** Deviance values changed as days in the reactivation group (i), the no-reactivation group (ii), the full-practice group (iii), the new reactivation group, and the new no-reactivation group (iv) under the same fitting method as Amar-Halpert et al. (2017) (a) and the psignifit4 fitting method (b). Solid triangles and hollow circles represented mean and individual values, respectively. Error bars indicated ± 1 standard error of the mean.

both fitting methods indicated the same observer (R23) in the reactivation group with poor goodness-of-fit on day 1 in the chi-square test. After excluding this observer's deviance values (which were retained in Fig. 6(i)), we performed a statistical analysis of the deviance values across days within each group. A one-way repeated measures ANOVA showed that, in the reactivation group, deviance values decreased significantly from day 1 (pre-test) to day 8 (post-test2) under the same fitting method (Fig. 6a(i), $p = 0.02$), but not under the psignifit4 fitting method (Fig. 6b(i), $p = 0.10$). In the no-reactivation group, the main effect of days is not significant (Fig. 6a(ii), $p = 0.35$; psignifit4: Fig. 6b (ii), $p = 0.65$). In the full-practice group, the main effect of days is also not significant (Fig. 6a(iii), $p = 0.35$; psignifit4: Fig. 6b(iii), $p = 0.36$). Additionally, paired samples t-tests indicated no significant change in deviance values from day 1 to day 5 for both the new reactivation and no-reactivation groups (Fig. 6a(iv), $ps > 0.1$; psignifit4: Fig. 6b(iv), $ps > 0.2$). These findings indicated that goodness-of-fit did not change across days in most groups, reflecting the stability of the fitting models.

Besides, both fitting methods indicated a significantly smaller deviance in the control experimental groups compared to the main experimental groups. Specifically, classical independent samples t-tests showed the new reactivation group had significantly lower deviance values compared to the reactivation group on day 1 ($p = 0.002$; psignifit4: $p < 0.001$) and day 5 ($p = 0.002$; psignifit4: $p < 0.001$). Similarly, the new no-reactivation group had significantly lower deviance values than the no-reactivation group on day 1 ($p = 0.005$; psignifit4: $p = 0.004$) and day 5 ($p < 0.001$; psignifit4: $p < 0.001$). These results suggested that our control experiment with the modified constant stimuli method produced better goodness of fit.

3.5. *Performance on the fixation task*

It is speculated that the large improvement of TDT in the no-reactivation group might be related to a strategy change in this group, where the observers shifted their focus of attention away from the fixation task towards the eccentric texture target discrimination task, thereby there might be a reduction in correctly reported fixation targets from the pre-test to the post-test. To answer this question, we analyzed the performance of the fixation task from the first day to the last day for each group. A one-way repeated measures ANOVA showed that the accuracies of the fixation task exhibited a significant increase from day 1 (pre-test) to the other four days (reactivation group: Fig. 7a, $ps < 0.001$; no-reactivation group: Fig. 7b, $ps < 0.01$). In the full-practice group, the accuracies increased significantly from day 1 (pre-test) to the last three days (Fig. 7c, $ps < 0.001$), and the accuracies on the second day were significantly lower than those on the last day (Fig. 7c, $p < 0.001$). Paired samples t-tests indicated that accuracies of the fixation task showed a significant increase from day 1 to day 5 for both the new reactivation and new no-reactivation groups (Fig. 7d, $ps < 0.05$). These results demonstrated a consistent increase rather than a reduction in correctly reported fixation tasks as the training progressed in all groups, thereby ruling out the possibility that the observed improvements, especially in the no-reactivation group and the new no-reactivation group, were due to a shift in performance strategy in favor of the eccentric task.

4. Discussion

In this study, we attempted to replicate the study of Amar-Halpert et al. (2017) using a larger number of observers and an improved experimental design. We did observe significant improvement in the reactivation group and the full-practice group as Amar-Halpert et al. (2017) showed. However, these improvements were comparable to that of the no-reactivation group. Moreover, further practice of the TDT task for an additional three days, the reactivation and no-reactivation groups showed additional significant improvements. After improving the schemes of the constant stimuli method, we still observed that improvements brought by reactivation and no-reactivation were

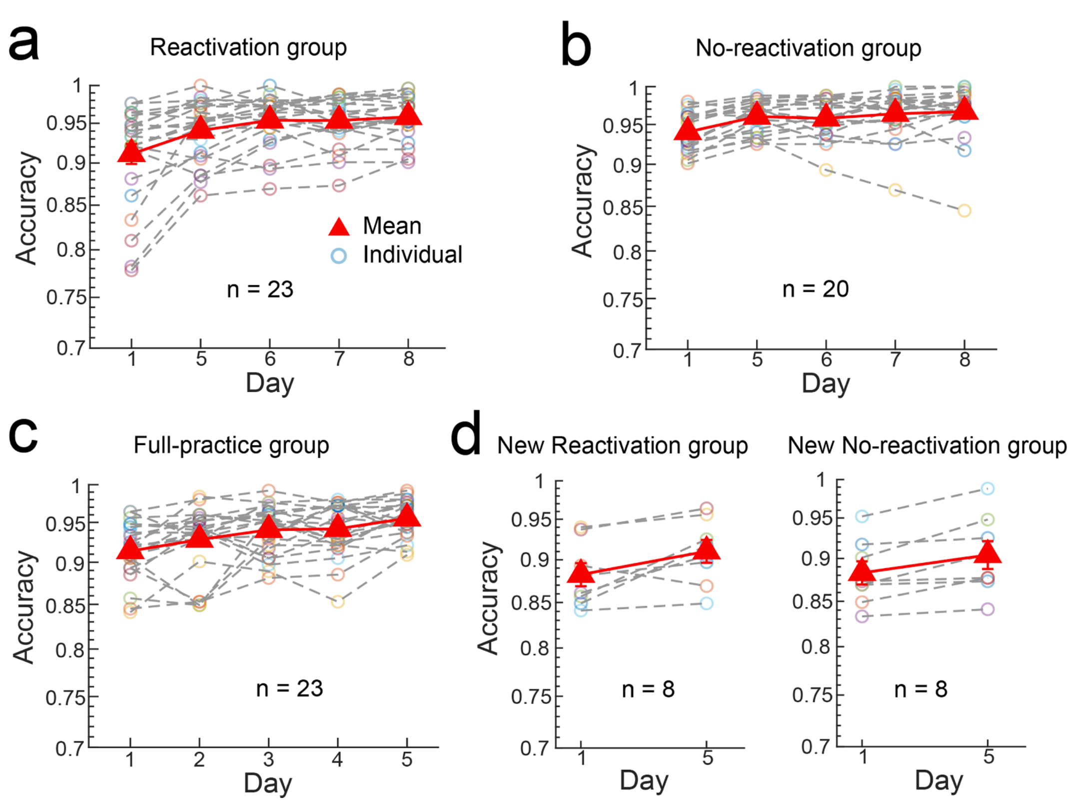

Fig. 7. Performance on the fixation task across days in each group. **a-d.** The accuracy of the fixation task changed as days in the reactivation group (a), the no-reactivation group (b), the full-practice group (c), the new reactivation group, and the new no-reactivation group (d). Solid triangles and hollow circles represented mean and individual values, respectively. Error bars indicated ± 1 standard error of the mean.

comparable. Our results suggested that the improvement in the TDT task associated with brief reactivation (5 trials per day over 3 days) did not surpass that of the no-reactivation group with no exposure to the TDT task over the same time period. Therefore, brief memory reactivation may not significantly contribute to the improvement of perceptual learning.

Both our main experiment and control experiment demonstrated that reactivation did not yield additional gains in learning improvement compared to a no-reactivation condition. The primary inconsistency between our study and Amar-Halpert et al. (2017) lies in the no-reactivation condition. In our no-reactivation group ($n = 23$), the pre-test and post-test were conducted three days apart, which aligned with the training schedules of the reactivation group, and we still found significant learning improvement. In contrast, Amar-Halpert et al. (2017) employed the improvement of two consecutive days (from day 1 to day 2) in the full-practice group ($n = 12$) as a no-reactivation control condition and found the improvement is insignificant (MPI $= 2.9 \pm 5.8$ %). This setting may be inappropriate as a no-reactivation control condition, given that the time interval from pre-test to post-test in the reactivation group was four days, whereas it was one day in this setting. The time interval may have a significant influence on the size of the retest effect (Scharfen et al., 2018). In addition, contrary to their finding, the learning progress from day 1 to day 2 in our full-practice group was significant (MPI $= 12.5 \pm 4.2$ %, using the same fitting method). On the other hand, it is noteworthy that Amar-Halpert et al. (2017) arranged a no-reactivation group ($n = 7$) in which the pre-test and post-test were spaced over nine days to align with their far-threshold reactivation condition, and they reported no progress in this no-reactivation group (with a threshold decrease of 7.6 ± 3.3 ms). The spaced time may be excessively long and could lead to memory decay, potentially explaining the lack of progress (F. Wang et al., 2016). With a relatively large sample size, our no-reactivation group followed a training schedule that was consistent with those of the reactivation and full-practice group and showed significant improvement that is comparable to that observed in the reactivation group. The improvement of the no-reactivation group can be attributed to significant retest effects or fast learning occurring in session, consistent with the widely observed significant retest effects in various tasks as shown in *meta*-analytic evidence (Scharfen, Jansen, & Holling, 2018; Scharfen, Peters, & Holling, 2018).

In phase I of our main experiment, the improvements in the full-practice group were comparable to those of the reactivation group, consistent with the results of Amar-Halpert et al. (2017). Recently the same group used fMRI to reveal the neural mechanisms of reactivation and standard repetition-based learning (Kondat et al., 2024). Again they demonstrated comparable improvements in a group with brief memory reactivations ($n = 20$) and a group with full practice ($n = 20$). Unfortunately, they did not arrange a no-reactivation group in this study. One might argue that the results from between-group comparisons were affected to some extent by individual differences (Chua et al., 2022; Dale et al., 2021; Yang et al., 2020). Hence, the within-group design of our main experiment, by having the reactivation and no-reactivation groups continue to practice the same task for an additional three days after the post-test on day 5, might overcome the limitations of between-group comparison. The total improvement of the two phases in both groups was significantly greater than the improvements achieved in Phase 1, demonstrating that the fullpractice effect was significantly superior to the effects of brief memory reactivation or simple test–retest. These results were consistent with the findings of Chen and de Beeck (2021), which demonstrated that the reactivation-induced improvement is significantly less than that achieved through full practice in more complex visual object learning. We speculated that the subtle improvements in our full-practice group could be partially attributed to the experimental design: most observers in this group experienced consecutive 5 days of full practice while the observers in the other two groups underwent spaced practice during the two phases of operation (~8 days). Time interval has been proven to play a substantial role in learning and memory, as spaced learning is reported to be more effective than massed learning (Raviv et al., 2022; Smolen et al., 2016; Vlach et al., 2008).

We initially employed the same constant stimuli method as described by Amar-Halpert et al. (2017), comprising 14 different SOAs, with each SOA containing only 18 trials. This method involved an excessive number of stimulus levels and an insufficient number of trials per stimulus level, potentially rendering the experimental data susceptible to the non-stationarity in observer behavior, which can lead to overdispersion in the data (Blackwell, 1952). The data overdispersion may signify a high level of uncertainty in threshold fitting (Schütt et al., 2016). In addition to the same fitting method as Amar-Halpert et al. (2017), we also provided the psignifit4 fitting method (Schütt et al., 2016), which allows accurate Bayesian estimation of psychometric functions for (potentially) overdispersed data. Though different fitting methods might induce varying threshold estimates (Kingdom & Prins, 2010; Manning et al., 2018), the two fitting methods draw consistent conclusions in our study. More importantly, we refined the constant stimuli method with fewer stimulus levels and more trials per level, which resulted in better fitting quality and possibly more accurate threshold estimation, and we still observed comparable improvements brought by reactivation and no-reactivation groups. It is worth pointing out that all groups in our study showed larger finger errors/lapse rate values on day 1. This issue raised an important question in perceptual learning research: the estimate of pre-test performance is so crucial to the quantification of perceptual learning and its transfer (Zhang & Yu, 2018). In particular, the estimation of pre-test performance should take into account the steep improvement that takes place in the first fifty or so trials, which may include learning that is not perceptual, but cognitive or procedural, such as how well the observer understands the task, instructions, response mappings, etc., as well as how to fixate and focus on the stimulus (Fahle et al., 1995; Karni & Sagi, 1993; Westheimer, 2001).

Whether reconsolidation following reactivation is a general property of all types of memory remains controversial. Studies with rodent models showed that although Pavlovian memory reconsolidation has been widely demonstrated, instrumental memory reconsolidation is still debated (Piva et al., 2020). Early investigations suggested that instrumental memories did not undergo reconsolidation (Brown et al., 2008; Hernandez & Kelley, 2004; Mierzejewski et al., 2009), while subsequent research indicated that these memories are just more resistant to destabilization and reconsolidation compared to Pavlovian memories (Exton-McGuinness et al., 2014; Tedesco et al., 2014). Likewise, human studies have demonstrated the reconsolidation of aversive and appetitive memory, as well as procedural memory related to motor skill tasks (Fan et al., 2020; Schwabe et al., 2014; Silva & Soares, 2018), whereas whether declarative memory could undergo reconsolidation is under debate (Chan & LaPaglia, 2013; Forcato et al., 2007; Hardwicke et al., 2016; Klingmüller et al., 2017). These discrepancies may be due to that memory involving more diverse and complex cortical circuits is more difficult to be modified via a reconsolidation process, as suggested by the fact that Pavlovian-conditioning fear memory involves relatively simple neural circuits centered at the amygdala, while declarative memories appear to involve broadly distributed neural circuits centered at the prefrontal cortex (Kim et al., 2021). Amar-Halpert et al. (2017) first reported that visual perceptual learning can be enhanced by reactivation and reconsolidation, but our results indicated that brief memory reactivation may not improve visual perception. Given the growing evidence supporting that visual perceptual learning operates at a conceptual level involving high-order cortical areas (Wang et al., 2016; Zhang et al., 2010), we speculate that this type of memory may not be easy to experience post-retrieval reconsolidation.

Several possible limitations to this study warrant discussion. First, in perceptual learning studies, the threshold is usually measured with the method of constant stimuli (Ahissar & Hochstein, 1997; Karni & Sagi, 1991), or adaptive procedures like the staircase procedure (Dosher & Lu, 1998; Xiao et al., 2008; Yu et al., 2004). Previous studies have shown that the exact conditions of measurement play an important role in learning and transfer (Manning et al., 2018; Xiong et al., 2016; Zhang & Yu, 2018). Further evidence is needed to determine whether the current results are specific to the particular psychophysical method. Second, although effective foveal fixation was shown as the high accuracy in the central tumbling T/L task, the observers could shift (without being aware of it) their gaze towards the trained quadrant with the TDT target by 1 to 2 degrees to gain resolution of the texture elements and reduce crowding. Further exploration combining eye tracking could be used to determine any changes in fixation behavior across practice sessions and to test whether the tumbling T/L task is completed by extra-foveal vision. Third, although our findings of the null effect of reactivation resonate with previous research in visual and other domains (Chalkia et al., 2021; Chen & de Beeck, 2021; Luyten & Beckers, 2017), the positive findings of Amar-Halpert et al. (2017) are supported by reactivation effects in related paradigms like orientation detection (Bang et al., 2018). In the domain of motor learning, length of reactivation was identified as a crucial boundary condition determining whether human motor memories can undergo reconsolidation (de Beukelaar et al., 2014). Similar boundary conditions have not been reported in visual perceptual learning, more replications and attempts are needed to confirm the existence of the reconsolidation phenomenon in the field of vision science.

CRediT authorship contribution statement

Jun-Ping Zhu: Writing – review & editing, Writing – original draft, Methodology, Investigation, Formal analysis, Data curation. **Jun-Yun Zhang:** Writing – review & editing, Writing – original draft, Visualization, Validation, Supervision, Software, Resources, Project administration, Methodology, Investigation, Funding acquisition, Formal analysis, Data curation, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This research was supported by the Natural Science Foundation of China grant 32371079 (JYZ; Beijing, China). We thank Cong Yu for his insightful comments and discussions. We thank CaiXia Chen and KunYun Wang for helping with the data analysis.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.visres.2025.108543.

Data availability

Data will be made available on request.

References

Ahissar, M., & Hochstein, S. (1997). Task difficulty and the specificity of perceptual learning. *Nature, 387*(6631), 401–406. https://doi.org/10.1038/387401a0

Amar-Halpert, R., Laor-Maayany, R., Nemni, S., Rosenblatt, J. D., & Censor, N. (2017). Memory reactivation improves visual perception. *Nature Neuroscience, 20*(10), Article 1325-+. https://doi.org/10.1038/nn.4629

Auber, A., Tedesco, V., Jones, C. E., Monfils, M. H., & Chiamulera, C. (2013). Post-retrieval extinction as reconsolidation interference: Methodological issues or boundary conditions? *Psychopharmacology (Berl), 226*(4), 631–647. https://doi.org/10.1007/s00213-013-3004-1

Bang, J. W., Shibata, K., Frank, S. M., Walsh, E. G., Greenlee, M. W., Watanabe, T., & Sasaki, Y. (2018). Consolidation and reconsolidation share behavioural and neurochemical mechanisms. *Nature Human Behaviour, 2*(7), 507–513. https://doi.org/10.1038/s41562-018-0366-8

Blackwell, H. R. (1952). Studies of Psychophysical Methods for Measuring Visual Thresholds*†. *Journal of the Optical Society of America, 42*(9), 606–616. https://doi.org/10.1364/JOSA.42.000606

Brown, T. E., Lee, B. R., & Sorg, B. A. (2008). The NMDA antagonist MK-801 disrupts reconsolidation of a cocaine-associated memory for conditioned place preference but not for self-administration in rats. *Learning & Memory, 15*(12).

Chalkia, A., Van Oudenhove, L., & Beckers, T. (2021). Preventing the return of fear in humans using reconsolidation update mechanisms: A verification report of Schiller et al. (2010) (vol 129, pg 510, 2020). *Cortex*, *139*, 309-309. doi: 10.1016/j.cortex.2021.03.010.

Chan, J. C. K., & LaPaglia, J. A. (2013). Impairing existing declarative memory in humans by disrupting reconsolidation [Article]. *Proceedings of the National Academy of Sciences of the United States of America, 110*(23), 9309–9313. https://doi.org/10.1073/pnas.1218472110

Chen, C.-Y., & de Beeck, H. O. (2021). Perceptual Learning with Complex Objects: A Comparison between Full-Practice Training and Memory Reactivation [Article]. *Eneuro*, *8*(2), Article 0008-19.2021. doi: 10.1523/eneuro.0008-19.2021.

Chua, S. Y. P., Rentzelas, P., Frangou, P., Kourtzi, Z., Lintern, M., & Mavritsaki, E. (2022). Cultural differences in visual perceptual learning. *International Journal of Psychology, 57*(3), 377–386. https://doi.org/10.1002/ijop.12824

Dale, G., Cochrane, A., & Green, C. S. (2021). Individual difference predictors of learning and generalization in perceptual learning. *Atten Percept Psychophys, 83*(5), 2241–2255. https://doi.org/10.3758/s13414-021-02268-3

Das, A., Tadin, D., & Huxlin, K. R. (2014). Beyond blindsight: Properties of visual relearning in cortically blind fields. *The Journal of Neuroscience, 34*(35), 11652–11664. https://doi.org/10.1523/jneurosci.1076-14.2014

Dayan, E., Laor-Maayany, R., & Censor, N. (2016). Reward disrupts reactivated human skill memory. *Scientific Reports, 6*(1). https://doi.org/10.1038/srep28270

de Beukelaar, T. T., Woolley, D. G., & Wenderoth, N. (2014). Gone for 60 seconds: Reactivation length determines motor memory degradation during reconsolidation. *Cortex, 59*, 138–145. https://doi.org/10.1016/j.cortex.2014.07.008

Dosher, B. A., & Lu, Z. L. (1998). Perceptual learning reflects external noise filtering and internal noise reduction through channel reweighting. *Proceedings of the National Academy of Sciences of the United States of America, 95*(23), 13988–13993. https://doi.org/10.1073/pnas.95.23.13988

Exton-McGuinness, M. T. J., Patton, R. C., Sacco, L. B., & Lee, J. L. C. (2014). Reconsolidation of a well-learned instrumental memory [Article]. *Learning & Memory, 21*(9), 468–477. https://doi.org/10.1101/lm.035543.114

Fahle, M., Edelman, S., & Poggio, T. (1995). Fast perceptual learning in hyperacuity. *Vision Research, 35*(21), 3003–3013. https://www.ncbi.nlm.nih.gov/pubmed/8533337.

Fan, C., Cheng, Y., Gou, H., Liu, C., Deng, S., Liu, C., Chen, X., Bu, J., & Zhang, X. (2020). Neuroimaging and intervening in memory reconsolidation of human drug addiction [Review]. *Science China-Information Sciences, 63*(7), Article 170103. https://doi.org/10.1007/s11432-019-2847-8

Forcato, C., Burgos, V. L., Argibay, P. F., Molina, V. A., Pedreira, M. E., & Maldonado, H. (2007). Reconsolidation of declarative memory in humans [Article]. *Learning & Memory, 14*(4), 295–303. https://doi.org/10.1101/lm.486107

Hardwicke, T. E., Taqi, M., & Shanks, D. R. (2016). Postretrieval new learning does not reliably induce human memory updating via reconsolidation. *Proceedings of the National Academy of Sciences, 113*(19), Article 201601440.

Haynes, J. D., Gallagher, M., Culling, J. F., & Freeman, T. C. A. (2024). The precision of signals encoding active self-movement [Article]. *Journal of Neurophysiology, 132*(2), 389–402. https://doi.org/10.1152/jn.00370.2023

Hernandez, P., & Kelley, A. (2004). Long-term memory for instrumental responses does not undergo protein synthesis-dependent reconsolidation upon retrieval. *Learning & Memory, 11*(6), 748–754.

Herpich, F., Melnick, M. D., Agosta, S., Huxlin, K. R., Tadin, D., & Battelli, L. (2019). Boosting Learning Efficacy with Noninvasive Brain Stimulation in Intact and Brain-Damaged Humans. *Journal of Neuroscience, 39*(28), 5551–5561. https://doi.org/10.1523/JNEUROSCI.3248-18.2019

Herszage, J., & Censor, N. (2017). Memory Reactivation Enables Long-Term Prevention of Interference. *Current Biology, 27*(10), Article 1529-+. https://doi.org/10.1016/j.cub.2017.04.025

Herszage, J., Sharon, H., & Censor, N. (2021). Reactivation-induced motor skill learning. *Proceedings of the National Academy of Sciences, 118*(23). https://doi.org/10.1073/pnas.2102242118

Hietanen, J. O., Syrjamaki, A. H., & Hietanen, J. K. (2022). Perception of eye contact, self-referential thinking and age [Article]. *Consciousness and Cognition, 106*, Article 103435. https://doi.org/10.1016/j.concog.2022.103435

Huang, Z. B., Niu, Z. M., & Li, S. (2023). Reactivation-induced memory integration prevents proactive interference in perceptual learning. *Journal of Vision, 23*(5). https://doi.org/10.1167/jov.23.5.1

Huxlin, K. R., Martin, T., Kelly, K., Riley, M., Friedman, D. I., Burgin, W. S., & Hayhoe, M. (2009). Perceptual Relearning of Complex Visual Motion after V1 Damage in Humans. *Journal of Neuroscience, 29*(13), 3981–3991. https://doi.org/10.1523/jneurosci.4882-08.2009

Jeter, P. E., Dosher, B. A., Liu, S. H., & Lu, Z. L. (2010). Specificity of perceptual learning increases with increased training. *Vision Research, 50*(19), 1928–1940. https://doi.org/10.1016/j.visres.2010.06.016

Johnson, V., Pramanik, S., & Shudde, R. (2022). Bayes factor functions for reporting outcomes of hypothesis tests. *arXiv*. doi: doi: 10.48550/arxiv.2210.00049.

Kantowitz, B. H., Iii, H. L. R., & Elmes, D. G. (2009). *Experimental psychology: Understanding psychological research* ((5th ed.). stress & health.).

Karni, A., & Sagi, D. (1991). Where Practice Makes Perfect in Texture-Discrimination - Evidence for Primary Visual-Cortex Plasticity. *Proceedings of the National Academy of Sciences of the United States of America, 88*(11), 4966–4970. https://doi.org/10.1073/pnas.88.11.4966

Karni, A., & Sagi, D. (1993). The time course of learning a visual skill. *Nature, 365*(6443), 250–252. http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=8371779.

Kim, G., Kwon, M., Kang, W., & Lee, S.-H. (2021). Is Reconsolidation a General Property of Memory? [Review]. *Frontiers in Human Neuroscience, 15*, Article 643106. https://doi.org/10.3389/fnhum.2021.643106

Kingdom, F. A. A., & Prins, N. (2010). *Psychophysics: A Practical Application.* Academic Press.

Kleiner, M., Brainard, D., & Pelli, D. (2007). What's new in Psychtoolbox-3? [Meeting Abstract]. *Perception, 36*, 14. <Go to ISI>://WOS:000250594600049.

Klingmüller, A., Caplan, J. B., & Sommer, T. (2017). *Intrusions in episodic memory: Reconsolidation or interference?* Cold Spring Harbor Lab(5).

Klorfeld-Auslender, S., Paz, Y., Shinder, I., Rosenblatt, J., Dinstein, I., & Censor, N. (2022). A distinct route for efficient learning and generalization in autism. *Current Biology*. https://doi.org/10.1016/j.cub.2022.05.059

Kondat, T., Tik, N., Sharon, H., Tavor, I., & Censor, N. (2024). Distinct Neural Plasticity Enhancing Visual Perception. *Journal of Neuroscience*, *44*(36). doi: ARTN e0301242024 10.1523/JNEUROSCI.0301-24.2024.

Lasagna, C. A., McLaughlin, M. M., Deng, W. Y., Whiting, E. L., & Tso, I. F. (2020). Deconstructing eye contact perception: Measuring perceptual precision and self-referential tendency using an online psychophysical eye contact detection task [Article]. *PLoS One, 15*(3), Article e0230258. https://doi.org/10.1371/journal.pone.0230258

Lee, J. L. C., Nader, K., & Schiller, D. (2017). An Update on Memory Reconsolidation Updating. *Trends in Cognitive Sciences, 21*(7), 531–545. https://doi.org/10.1016/j.tics.2017.04.006

Levi, D. M., & Polat, U. (1996). Neural plasticity in adults with amblyopia [Meeting Abstract]. *Investigative Ophthalmology & Visual Science, 37*(3), 3079. <Go to ISI>://WOS:A1996TX39703075.

Li, R. W., Klein, S. A., & Levi, D. M. (2008). Prolonged perceptual learning of positional acuity in adult amblyopia: Perceptual template retuning dynamics. *Journal of Neuroscience, 28*(52), 14223–14229. http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=19109504.

Lu, Z. L., & Dosher, B. A. (2022). Current directions in visual perceptual learning. *Nature Reviews Psychology, 1*(11), 654–668. https://doi.org/10.1038/s44159-022-00107-2

Luyten, L., & Beckers, T. (2017). A preregistered, direct replication attempt of the retrieval-extinction effect in cued fear conditioning in rats. *Neurobiology of Learning and Memory, 144*, 208–215. https://doi.org/10.1016/j.nlm.2017.07.014

Manning, C., Jones, P. R., Dekker, T. M., & Pellicano, E. (2018). Psychophysics with children: Investigating the effects of attentional lapses on threshold estimates. *Attention Perception & Psychophysics, 80*(5), 1311–1324. https://doi.org/10.3758/s13414-018-1510-2

Mierzejewski, P., Korkosz, A., Rogowski, A., Korkosz, I., Kostowski, W., & Scinska, A. (2009). Post-session injections of a protein synthesis inhibitor, cycloheximide do not alter saccharin self-administration. *Progress in Neuro-Psychopharmacology and Biological Psychiatry, 33*(2), 286–289.

Misanin, J. R., Miller, R. R., & Lewis, D. J. (1968). Retrograde Amnesia Produced by Electroconvulsive Shock after Reactivation of a Consolidated Memory Trace. [Article]. *Science, 160*(3827), 554-+. https://doi.org/10.1126/science.160.3827.554

Nader, K., Schafe, G. E., & Le Doux, J. E. (2000). Fear memories require protein synthesis in the amygdala for reconsolidation after retrieval [Article]. *Nature, 406*(6797), 722–726. https://doi.org/10.1038/35021052

Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: Transforming numbers into movies. *Spatial Vision, 10*(4), 437–442. https://doi.org/10.1163/156856897x00366

Piva, A., Pintori, N., Padovani, L., & Chiamulera, C. (2020). Protocols for instrumental memory reconsolidation in rodents: A methodological review [Article; Proceedings Paper]. *Journal of Neuroscience Methods, 342*, Article 108766. https://doi.org/10.1016/j.jneumeth.2020.108766

Plank, T., Rosengarth, K., Schmalhofer, C., Goldhacker, M., Brandl-Ruehle, S., & Greenlee, M. W. (2014). Perceptual learning in patients with macular degeneration. Article 1189 *Frontiers in Psychology, 5*. https://doi.org/10.3389/fpsyg.2014.01189.

Raviv, L., Lupyan, G., & Green, S. C. (2022). How variability shapes learning and generalization. *Trends in Cognitive Sciences, 26*(6), 462–483. https://doi.org/10.1016/j.tics.2022.03.007

Sagi, D. (2011). Perceptual learning in Vision Research. *Vision Research, 51*(13), 1552–1566. <Go to ISI>://WOS:000293549100009.

Scharfen, J., Jansen, K., & Holling, H. (2018). Retest effects in working memory capacity tests: A meta-analysis [Review]. *Psychonomic Bulletin & Review, 25*(6), 2175–2199. https://doi.org/10.3758/s13423-018-1461-6

Scharfen, J., Peters, J. M., & Holling, H. (2018). Retest effects in cognitive ability tests: A meta-analysis [Article]. *Intelligence, 67*, 44–66. https://doi.org/10.1016/j.intell.2018.01.003

Schneider, A. M., & Sherman, W. (1968). Amnesia: A function of the temporal relation of footshock to electroconvulsive shock. *Science, 159*(3811), 219-+. https://doi.org/10.1126/science.159.3811.219

Schrift, G., Dotan, D., & Censor, N. (2022). Brief memory reactivations induce learning in the numeric domain. *npj Science of Learning, 7*(1). https://doi.org/10.1038/s41539-022-00136-9

Schütt, H. H., Harmeling, S., Macke, J. H., & Wichmann, F. A. (2016). Painfree and accurate Bayesian estimation of psychometric functions for (potentially) overdispersed data. *Vision Research, 122*, 105–123. https://doi.org/10.1016/j.visres.2016.02.002

Schwabe, L., Nader, K., & Pruessner, J. C. (2014). Reconsolidation of Human Memory: Brain Mechanisms and Clinical Relevance [Review]. *Biological Psychiatry, 76*(4), 274–280. https://doi.org/10.1016/j.biopsych.2014.03.008

Shibata, K., Sasaki, Y., Bang, J. W., Walsh, E. G., Machizawa, M. G., Tamaki, M., Chang, L.-H., & Watanabe, T. (2017). Overlearning hyperstabilizes a skill by rapidly making neurochemical processing inhibitory-dominant. *Nature Neuroscience, 20*(3), 470–475. https://doi.org/10.1038/nn.4490

Silva, M. B., & Soares, A. B. (2018). Reconsolidation of human motor memory: From boundary conditions to behavioral interventions-How far are we from clinical applications? [Review]. *Behavioural Brain Research, 353*, 83–90. https://doi.org/10.1016/j.bbr.2018.07.003

Smolen, P., Zhang, Y. L., & Byrne, J. H. (2016). The right time to learn: Mechanisms and optimization of spaced learning. *Nature Reviews Neuroscience, 17*(2), 77–88. https://doi.org/10.1038/nrn.2015.18

Su, Y., Shi, Z., & Wachtler, T. (2024). A Bayesian observer model reveals a prior for natural daylights in hue perception [Article]. *Vision Research, 220*, Article 108406. https://doi.org/10.1016/j.visres.2024.108406

Tedesco, V., Mutti, A., Auber, A., & Chiamulera, C. (2014). Nicotine-seeking reinstatement is reduced by inhibition of instrumental memory reconsolidation [Article]. *Behavioural Pharmacology, 25*(8), 725–731. https://doi.org/10.1097/fbp.0000000000000088

Vlach, H. A., Sandhofer, C. M., & Kornell, N. (2008). The spacing effect in children's memory and category induction. *Cognition, 109*(1), 163–167. https://doi.org/10.1016/j.cognition.2008.07.013

Wagenmakers, E.-J., Love, J., Marsman, M., Jamil, T., Ly, A., Verhagen, J., Selker, R., Gronau, Q. F., Dropmann, D., Boutin, B., Meerhoff, F., Knight, P., Raj, A., van Kesteren, E.-J., van Doorn, J., Smira, M., Epskamp, S., Etz, A., Matzke, D., & Morey, R. D. (2018). Bayesian inference for psychology. Part II: Example applications with JASP [Article]. *Psychonomic Bulletin & Review, 25*(1), 58–76. https://doi.org/10.3758/s13423-017-1323-7

Walker, M. P., Brakefield, T., Hobson, J. A., & Stickgold, R. (2003). Dissociable stages of human memory consolidation and reconsolidation. *Nature, 425*(6958), 616–620. http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=14534587.

Wang, F., Huang, J., Lv, Y., Ma, X., Yang, B., Wang, E., Du, B., Li, W., & Song, Y. (2016). Predicting perceptual learning from higher-order cortical processing. *Neuroimage, 124*(Pt A), 682–692. https://doi.org/10.1016/j.neuroimage.2015.09.024

Wang, R., Cong, L. J., & Yu, C. (2013). The classical TDT perceptual learning is mostly temporal learning. *Journal of Vision, 13(5):9*, 1–9. http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=20844128.

Wang, R., Wang, J., Zhang, J. Y., Xie, X. Y., Yang, Y. X., Luo, S. H., Yu, C., & Li, W. (2016). Perceptual Learning at a Conceptual Level. *Journal of Neuroscience, 36*(7), 2238–2246. https://doi.org/10.1523/JNEUROSCI.2732-15.2016

Watanabe, T., & Sasaki, Y. (2015). Perceptual learning: Toward a comprehensive theory. *Annu Rev Psychol, 66*, 197–221. https://doi.org/10.1146/annurev-psych-010814-015214

Westheimer, G. (2001). Is peripheral visual acuity susceptible to perceptual learning in the adult? *Vision Research, 41*(1), 47–52. http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=11163615.

Wichmann, F. A., & Hill, N. J. (2001). The psychometric function: I. Fitting, sampling, and goodness of fit [Article]. *Perception and Psychophysics, 63*(8), 1293–1313. https://doi.org/10.3758/BF03194544

Xiao, L. Q., Zhang, J. Y., Wang, R., Klein, S. A., Levi, D. M., & Yu, C. (2008). Complete transfer of perceptual learning across retinal locations enabled by double training. *Current Biology, 18*(24), 1922–1926. http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=19062277.

Xiong, Y.-Z., Xie, X.-Y., & Yu, C. (2016). Location and direction specificity in motion direction learning associated with a single-level method of constant stimuli. *Vision Research, 119*, 9–15. https://doi.org/10.1016/j.visres.2015.11.005

Yang, J., Yan, F.-F., Chen, L., Xi, J., Fan, S., Zhang, P., Lu, Z.-L., & Huang, C.-B. (2020). General learning ability in perceptual learning. *Proceedings of the National Academy of Sciences of the United States of America, 117*(32), 19092–19100. https://doi.org/10.1073/pnas.2002903117

Yu, C., Klein, S. A., & Levi, D. M. (2004). Perceptual learning in contrast discrimination and the (minimal) role of context. *J Vis, 4*(3), 169–182. https://doi.org/10.1167/4.3.4

Zhang, J. Y., Cong, L. J., Klein, S. A., Levi, D. M., & Yu, C. (2014). Perceptual learning improves adult amblyopic vision through rule-based cognitive compensation. *Invest Ophthalmol Vis Sci, 55*(4), 2020–2030. https://doi.org/10.1167/iovs.13-13739

Zhang, J. Y., & Yu, C. (2018). Vernier learning with short- and long-staircase training and its transfer to a new location with double training. *J Vis, 18*(13), 8. https://doi.org/10.1167/18.13.8

Zhang, J. Y., Zhang, G. L., Xiao, L. Q., Klein, S. A., Levi, D. M., & Yu, C. (2010). Rule-based learning explains visual perceptual learning and its specificity and transfer. *Journal of Neuroscience, 30*(37), 12323–12328. https://doi.org/10.1523/JNEUROSCI.0704-10.2010